

GenCore version 5.1.3
Copyright (c) 1993 - 2002 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: November 28, 2002, 02:32:11 ; Search time 94 seconds
(without alignments)
7399.393 Million cell updates/sec

Title: US-09-642-034-4
Perfect score: 2268
Sequence: 1 atggcgagggaattatctgt.....tggttcgtataaaatttctag 2268

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 441362 seqs, 15338381 residues

Total number of hits satisfying chosen parameters: 882724

Minimum PR seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Issued Patents NA: *
1: /cgn2_6/ptodata/2/ina/5A-COMB.seq: *
2: /cgn2_6/ptodata/2/ina/5B-COMB.seq: *
3: /cgn2_6/ptodata/2/ina/6A-COMB.seq: *
4: /cgn2_6/ptodata/2/ina/6B-COMB.seq: *
5: /cgn2_6/ptodata/2/ina/PCTUS-COMB.seq: *
6: /cgn2_6/ptodata/2/ina/backfiles1.seq: *

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	DB ID	Description
1	1263	55.7	1310	1	US-08-311-023-1
2	1263	55.7	2404	1	US-08-311-023-3
3	59.8	2.6	285	4	US-08-990-571-64
4	59.8	2.6	285	4	US-09-528-784A-64
5	59.8	2.6	342	4	US-08-990-571-65
6	59.8	2.6	342	4	US-09-528-784A-65
7	59.8	2.6	351	4	US-08-990-571-60
8	59.8	2.6	351	4	US-09-528-784A-60
9	59.8	2.6	356	4	US-08-990-571-63
10	59.8	2.6	356	4	US-09-528-784A-63
11	59.8	2.6	358	4	US-08-990-571-58
12	59.8	2.6	358	4	US-09-528-784A-58
13	59.8	2.6	363	4	US-08-990-571-66
14	59.8	2.6	363	4	US-09-528-784A-66
15	59.8	2.6	363	4	US-08-990-571-67
16	59.8	2.6	363	4	US-09-528-784A-67
17	59.8	2.6	409	4	US-08-990-571-59
18	59.8	2.6	409	4	US-09-528-784A-59
19	59.8	2.6	410	4	US-08-990-571-61
20	59.8	2.6	410	4	US-09-528-784A-61
21	59.8	2.6	416	4	US-08-990-571-62
22	59.8	2.6	416	4	US-09-528-784A-62
23	59.8	2.6	1271	4	US-08-845-258-5
24	59.8	2.6	1271	4	US-08-845-258-42
25	59.8	2.6	1271	4	US-08-990-571-5
26	59.8	2.6	1271	4	US-08-990-571-42
27	59.8	2.6	1271	4	US-08-723-142A-5

C 28	59.8	2.6	1271	4	US-08-723-142A-42	Sequence 42, Appl
C 29	59.8	2.6	1271	4	US-09-528-784A-5	Sequence 5, Appl
C 30	59.8	2.6	1271	4	US-09-528-784A-42	Sequence 42, Appl
C 31	59.8	2.6	1821	4	US-08-845-258-6	Sequence 6, Appl
C 32	59.8	2.6	1821	4	US-08-990-571-6	Sequence 6, Appl
C 33	59.8	2.6	1821	4	US-08-723-142A-6	Sequence 6, Appl
C 34	59.8	2.6	1821	4	US-09-528-784A-6	Sequence 6, Appl
C 35	59.8	2.6	4223	4	US-08-845-258-7	Sequence 7, Appl
C 36	59.8	2.6	4223	4	US-08-990-571-7	Sequence 7, Appl
C 37	59.8	2.6	4223	4	US-08-845-258-45	Sequence 45, Appl
C 38	59.8	2.6	4223	4	US-08-990-571-45	Sequence 45, Appl
C 39	59.8	2.6	4223	4	US-08-723-142A-7	Sequence 7, Appl
C 40	59.8	2.6	4223	4	US-08-723-142A-45	Sequence 45, Appl
C 41	59.8	2.6	4223	4	US-09-528-784A-7	Sequence 7, Appl
C 42	59.8	2.6	4223	4	US-09-528-784A-45	Sequence 45, Appl
C 43	57.2	2.5	18596	4	US-09-318-448-11	Sequence 11, Appl
C 44	56.2	2.5	5552	3	US-08-155-888-1	Sequence 1, Appl
C 45	55.6	2.5	377	2	US-08-332-766A-1	Sequence 1, Appl

ALIGNMENTS

RESULT 1
US-08-311-023-1
; Sequence 1, Application US/08311023
; Patent No. 5693465
; GENERAL INFORMATION:
; APPLICANT: MANNING, David Lockwood
; APPLICANT: NICHOLSON, Robert Ian
; APPLICANT: GEE, Julia Margaret
; APPLICANT: GREEN, Christopher Douglas
; TITLE OF INVENTION: METHODS FOR PREDICTING THE BEHAVIOUR OF
; TITLE OF INVENTION: BREAST TUMOURS
; NUMBER OF SEQUENCES: 3
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Young & Thompson
; STREET: 745 South 23rd Street
; CITY: Arlington
; STATE: VA
; COUNTRY: USA
; ZIP: 22202

COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25 (EPO)
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/311,023
; FILING DATE: 22-SEP-1994
; ATTORNEY/AGENT INFORMATION:
; NAME: PATCH, Andrew J.
; REGISTRATION NUMBER: Reg. No. 5693465 32.925
; REFERENCE/DOCKET NUMBER: WCM.56
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 703/521-2297
; TELEFAX: 703/685-0573
; TELEX: 248425
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1310 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: CDNA
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 1..1293
; US-08-311-023-1

Query Match 55.7%; Score 1263; DB 1; Length 1310;
Best Local Similarity 99.8%; Pred. No. 0;
Matches 1285; Conservative 0; Mismatches 0; Indels 2; Gaps 2;

```
QY 982 GGTATTTATAGCCATTTCATCATCAGTTTCCTGCTCTCTGCTGGGGGTATCTTATAGTGCCT 1041
Db 1 GGTATTTATAGCCATTTCATCATCAGTTTCCTGCTCTCTGCTGGGGGTATCTTATAGTGCCT 60
QY 1042 CTCATGAATCGGGGTGTTTTTCAAAATTCCTCTGAGTTTCCCTGTGTGGCAGCTGGCGGTGGG 1101
Db 61 CTCATGAATCGGGGTGTTTTTCAAAATTCCTCTGAGTTTCCCTGTGTGGCAGCTGGCGGTGGG 120
QY 1102 ACTTTTCAGTGGTGATGCTTTTTTACACCTTCTTCCACATCTCTCAATCTCAATGCAAGTTCACGACAT 1161
Db 121 ACTTTTCAGTGGTGATGCTTTTTTACACCTTCTTCCACATCTCTCAATCTCAATGCAAGTTCACGACAT 180
QY 1162 AGTCATAGCCATGAAGAACACAGCAATGAAATGAAAGAGAGGACATTTTTCAGTCATCTG 1221
Db 181 AGTCATAGCCATGAAGAACACAGCAATGAAATGAAAGAGAGGACATTTTTCAGTCATCTG 240
QY 1222 TCCTTCTCAAAACATAGAAGAAGTGCCTTATTTGATTCACAGTGGGAAGGCTTAACAGCT 1281
Db 241 TCCTTCTCAAAACATAGAAGAAGTGCCTTATTTGATTCACAGTGGGAAGGCTTAACAGCT 300
QY 1282 CTAGGAGGCTCTATTTCATGTTCTTGTGGAACATCTCTCATGATGATCAAAACAAATTT 1341
Db 301 CTAGGAGGCTCTATTTCATGTTCTTGTGGAACATCTCTCATGATGATCAAAACAAATTT 360
QY 1342 AAAGATAAGAAAGAAAGATCAGAAGAAACCTGAAAATGATGATGTGGAGATTAAAG 1401
Db 361 AAAGATAAGAAAGAAAGATCAGAAGAAACCTGAAAATGATGATGTGGAGATTAAAG 420
QY 1402 AAGCAGTTGTCACAGTATGATCTCAACTTTCACAAATCAGGAGAAATGATGATCAGAT 1461
Db 421 AAGCAGTTGTCACAGTATGATCTCAACTTTCACAAATCAGGAGAAATGATGATCAGAT 480
QY 1462 GATCGAATGAAGGCTATTTACGAGCAGACTCACAGAGCCCTCCCACTTTGATTTCTCAG 1521
Db 481 GATCGAATGAAGGCTATTTACGAGCAGACTCACAGAGCCCTCCCACTTTGATTTCTCAG 540
QY 1522 CAGCTTCACGCTTGGGAAGAGAGAGGTTCATGATAGCTCATGCTCATGCTCATGCTCATGCT 1581
Db 541 CAGCTTCACGCTTGGGAAGAGAGAGGTTCATGATAGCTCATGCTCATGCTCATGCTCATGCT 600
QY 1582 TACAAATGAATATGTACCCAGAGGCTGCAAGAAATGAATGCAATGCAATTTTCCAGCATACA 1641
Db 601 TACAAATGAATATGTACCCAGAGGCTGCAAGAAATGAATGCAATTTTCCAGCATACA 660
QY 1642 CTCGCCCATGACAGATCTCATTTACCCAGCATGATGATGATGATGATGATGATGATGATGAT 1701
Db 661 CTCGCCCATGACAGATCTCATTTACCCAGCATGATGATGATGATGATGATGATGATGATGAT 720
QY 1702 CACCAACCAACCAACCAACCACTCTCAGCTCAGCTCAGCTCAGCTCAGCTCAGCTCAGCTCAG 1761
Db 721 CACCAACCAACCAACCAACCACTCTCAGCTCAGCTCAGCTCAGCTCAGCTCAGCTCAGCTCAG 780
QY 1762 AAGATGCGGGCTGCGCCACTTTGGCTGGATGGTGAATAGGCTGATGGCTGATGGCTGATGG 1821
Db 781 AAGATGCGGGCTGCGCCACTTTGGCTGGATGGTGAATAGGCTGATGGCTGATGGCTGATGG 840
QY 1822 TTCAGCGATGCGCTACCAATTTGCTGCTGCTTTTACTGAAGGCTTATCAAGTGGTTTAAAT 1881
Db 841 TTCAGCGATGCGCTACCAATTTGCTGCTGCTTTTACTGAAGGCTTATCAAGTGGTTTAAAT 900
QY 1882 ACTTCTGTTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1941
Db 901 ACTTCTGTTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 960
QY 1942 CTAAGGCTGGCATGACCTTTAAGCAGGCTGCTCTTTAATATGCAATGCTGCAAGTGGT 2001
Db 961 CTAAGGCTGGCATGACCTTTAAGCAGGCTGCTCTTTAATATGCAATGCTGCAAGTGGT 1020
QY 2002 GCGTATCTTGGAAATGCAACAGGAATTTTCATTTGCTGCTATTTGCTGCAAAATGCTTTCATG 2061
Db 1021 GCGTATCTTGGAAATGCAACAGGAATTTTCATTTGCTGCTATTTGCTGCAAAATGCTTTCATG 1080
```

```
QY 2062 TGGATATTTGCACTTACTGCTGCTTATTTCATGATGTTGCTGCTGCTGCTGCTGCTGCT 2121
Db 1081 TGGATATTTGCACTTACTGCTGCTTATTTCATGATGTTGCTGCTGCTGCTGCTGCTGCT 1140
QY 2122 GAAATGCTGCACAATGATGCTAGTGACCATGGATGTAGCCGCTGGGGGTATTTCTTTTAA 2181
Db 1141 GAAATGCTGCACAATGATGCTAGTGACCATGGATGTAGCCGCTGGGGGTATTTCTTTTAA 1200
QY 2182 CAGAATGCTGGGATGCTTTTGGGTTTGGAAATTAATGTTACTTATTTCCATATTTGAACAT 2241
Db 1201 CAGAATGCTGGGATGCTTTTGGGTTTGGAAATTAATGTTACTTATTTCCATATTTGAACAT 1259
QY 2242 AAAATCGTGTTCGTATATAAATTTCTAG 2268
Db 1260 AAAATCGTGTTCGTATATAAATTTCTAG 1285
```

RESULT 2

```
US-08-311-023-3
; Sequence 3, Application US/08311023
; Patent No. 5693465
; GENERAL INFORMATION:
; APPLICANT: MANNING, David Lockwood
; APPLICANT: NICHOLSON, Robert Ian
; APPLICANT: GEE, Julia Margaret
; APPLICANT: GREEN, Christopher Douglas
; TITLE OF INVENTION: METHODS FOR PREDICTING THE BEHAVIOUR OF
; TITLE OF INVENTION: BREAST TUMOURS
; NUMBER OF SEQUENCES: 3
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Young & Thompson
; STREET: 745 South 23rd Street
; CITY: Arlington
; STATE: VA
; COUNTRY: USA
; ZIP: 22202
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.25 (EPO)
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/311,023
; FILING DATE: 22-SEP-1994
; ATTORNEY/AGENT INFORMATION:
; NAME: PATCH, Andrew J.
; REGISTRATION NUMBER: Reg. No. 5693465 32.925
; REFERENCE/DOCKET NUMBER: WCM.56
; TELEPHONE: 703/521-2297
; TELEFAX: 703/685-0573
; TELEX: 248425
; INFORMATION FOR SEQ ID NO: 3:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2404 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: cDNA
US-08-311-023-3
```

```
Query Match 55.7%; Score 1263; DB 1; Length 2404;
Best Local Similarity 99.8%; Pred. No. 0;
Matches 1285; Conservative 0; Mismatches 0; Indels 2; Gaps 2;
```

```
QY 982 GGTATTTATAGCCATTTCATCATCAGTTTCCTGCTCTCTGCTGGGGGTATCTTATAGTGCCT 1041
Db 1 GGTATTTATAGCCATTTCATCATCAGTTTCCTGCTCTCTGCTGGGGGTATCTTATAGTGCCT 60
QY 1042 CTCATGAATCGGGGTGTTTTTCAAAATTCCTCTGAGTTTCCCTGAGTTTCTTGTGGCAGCTGGCGGTGGG 1101
Db 61 CTCATGAATCGGGGTGTTTTTCAAAATTCCTCTGAGTTTCCCTGAGTTTCTTGTGGCAGCTGGCGGTGGG 120
```

QY 1102 ACTTTGAGTGGTGATGCTTTTACACATTTCTTCCACATTTCTCATGCAATCAACAGCAT 1161
DB 121 ACTTTGAGTGGTGATGCTTTTACACATTTCTTCCACATTTCTCATGCAATCAACAGCAT 180
QY 1162 AGTCATAGCCATGAAGAACCAAGCAATCGAATGAAGAGAGACCACTTTTCAATCATCTG 1221
DB 181 AGTCATAGCCATGAAGAACCAAGCAATCGAATGAAGAGAGACCACTTTTCAATCATCTG 240
QY 1222 TCTTCTCAAAACATAGAAAGAGTGCCTATTTTGATTCCACGTGGAGAGTCTAAACAGCT 1281
DB 241 TCTTCTCAAAACATAGAAAGAGTGCCTATTTTGATTCCACGTGGAGAGTCTAAACAGCT 300
QY 1282 CTAGAGGCGCTGATTTTCATGTTTCTTGAACATGCTCTCACATTTGATCAAAACATTT 1341
DB 301 CTAGAGGCGCTGATTTTCATGTTTCTTGAACATGCTCTCACATTTGATCAAAACATTT 360
QY 1342 AAACATAGAGAAAGAAATCAAGAAACCTCAAAATCATGATGATGAGAGATTAAG 1401
DB 161 AAACATAGAGAAAGAAATCAAGAAACCTCAAAATCATGATGATGAGAGATTAAG 420
QY 1402 AAGCACTTTGTCCAAAGTATGAATCTCAACTTTTCAACAAATCAGGAGAACTAGATACAGAT 1461
DB 421 AAGCACTTTGTCCAAAGTATGAATCTCAACTTTTCAACAAATCAGGAGAACTAGATACAGAT 480
QY 1462 GATCGAATCAAGGCTATTTACGAGCAGACTCAAGAGCCCTCCCACTTTTGATTTCTCAG 1521
DB 481 GATCGAATCAAGGCTATTTACGAGCAGACTCAAGAGCCCTCCCACTTTTGATTTCTCAG 540
QY 1522 CAGCTCTCAGCTCTTGGAGAGAGAGGTCATGATAGCTCATGCTCATCACAGAGAGTC 1581
DB 541 CAGCTCTCAGCTCTTGGAGAGAGAGGTCATGATAGCTCATGCTCATCACAGAGAGTC 600
QY 1582 TACAATCAATATGTACCCAGAGGCTGCAAGATAATGCCATTTCACATTTCCACGATACA 1641
DB 601 TACAATCAATATGTACCCAGAGGCTGCAAGATAATGCCATTTCACATTTCCACGATACA 660
QY 1642 CTCGAGCAGTCAGAGCTCATTTACAGCAGACTCATGATGATGATGATGATGATGATGAT 1701
DB 661 CTCGAGCAGTCAGAGCTCATTTACAGCAGACTCATGATGATGATGATGATGATGATGATGAT 720
QY 1702 CACCAACCAAAACCAACCTCTCAGTCAGAGCAGGCTTACCTCGAGAGAGCTG 1761
DB 721 CACCAACCAAAACCAACCTCTCAGTCAGAGCAGGCTTACCTCGAGAGAGCTG 780
QY 1762 AAAGATGCGCGCTGCGCCACTTTGGCTGGATGATGATGATGATGATGATGATGATGATGAT 1821
DB 781 AAAGATGCGCGCTGCGCCACTTTGGCTGGATGATGATGATGATGATGATGATGATGATGAT 840
QY 1822 TTTACAGCATGCGCTAGCAATTTGCTGCTTTTACTGAAGGCTTATCAAGTGTGTTAACT 1881
DB 841 TTTACAGCATGCGCTAGCAATTTGCTGCTTTTACTGAAGGCTTATCAAGTGTGTTAACT 900
QY 1882 ACTTCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1941
DB 901 ACTTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 960
QY 1942 CTAAGGCTGCGCATGACCGTTAAGCAGGCTGCTCTTTATATGCTGCTGCTGCTGCTGCTGCT 2001
DB 961 CTAAGGCTGCGCATGACCGTTAAGCAGGCTGCTCTTTATATGCTGCTGCTGCTGCTGCTGCT 1020
QY 2002 GCGTATCTTGGATGCAACAGAAATTTTCATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 2061
DB 1021 GCGTATCTTGGATGCAACAGAAATTTTCATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1080
QY 2062 TGGATATTTGCACTTACTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 2121
DB 1081 TGGATATTTGCACTTACTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1140
QY 2122 GAAATGCTGCAATGATGCTAGTGACCATGAGATGAGCCGCTGGGGGATTTTCCTTTT 2181
DB 1141 GAAATGCTGCAATGATGCTAGTGACCATGAGATGAGCCGCTGGGGGATTTTCCTTTT 1200
QY 2182 CACAATCTCTGGGATGCTTTTGGGTTTGGAAATVATGTTACTTATTTTCATATTTGAAAT 2241

DB 1201 CAGATGCTGGGATGCTTTTGGGTTTGGAAATATGTTACTTA-TTCCATATTTGAACAT 1259
QY 2242 AAAATCGTCTTTTCGTATAAAATTTCTAG 2268
DB 1260 AAAATCGTCTTTTCGTATAAAATTTCTAG 1285

RESULT 3

US-08-990-571-64/c
; Sequence 64, Application US/08990571
; Patent No. 6214971
; GENERAL INFORMATION:
; APPLICANT: Reed, Steven G. et al.
; TITLE OF INVENTION: COMPOUNDS AND METHODS FOR THE DIAGNOSIS AND TREATMENT OF B.
; NUMBER OF SEQUENCES: 79
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: SEED AND BERRY
; STREET: 6300 Columbia Center, 701 Fifth Avenue
; CITY: Seattle
; STATE: Washington
; COUNTRY: USA
; ZIP: 98104
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/990,571
; FILING DATE: 11-DEC-1997
; CLASSIFICATION:
; ATTORNEY/AGENT INFORMATION:
; NAME: Makl, David J.
; REGISTRATION NUMBER: 31,392
; REFERENCE/DOCKET NUMBER: 210121.426C2
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (206) 622-4900
; TELEFAX: (206) 682-6031
; INFORMATION FOR SEQ ID NO: 64:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 285 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear

US-08-990-571-64

Query Match 2.6%; Score 59.8; DB 4; Length 285;
Best Local Similarity 65.2%; Pred. No. 7.3e-08;
Matches 88; Conservative 0; Mismatches 47; Indels 0; Gaps 0;

QY 268 ATCCATATACACCATGACCCAGCACCATCACTCAGACGAGCATCACTCAGACCATGAG 327
DB 192 AACCAAGTTCCACTAGGCCCCAGCAGCTTCACTAGGCCCCAGCTTCACTAGGCCCCAG 133
QY 328 COTCACTCAGACCATGAGCATCACTCAGACCCAGGAGCATCTCTGACCATGATCATCAC 387
DB 132 CTTCACTAGGCCCCAGCAGCTTCACTAGGCCCCAGCTTCACTAGGCCCCAGCTTCACT 73
QY 388 TCTCACCATAATCAT 402
DB 72 TAGGCCCCAGCTT 58

RESULT 4

US-09-528-784A-64/c
; Sequence 64, Application US/09528784A
; Patent No. 6451315
; GENERAL INFORMATION:
; APPLICANT: Reed, Steven G.
; APPLICANT: Iodes, Michael J.
; APPLICANT: Houghlon, Raymond L.
; APPLICANT: Sleath, Paul R.

RESULT 10
US-09-528-784A-63/C
; Sequence 63, Application US/09528784A
; Patent No. 6451315
; GENERAL INFORMATION:
; APPLICANT: Reed, Steven G.
; APPLICANT: Lodes, Michael J.
; APPLICANT: Houghton, Raymond L.
; APPLICANT: Sleath, Paul R.
; APPLICANT: McNeill, Patricia D.
; TITLE OF INVENTION: COMPOUNDS AND METHODS FOR THE DIAGNOSIS
; TITLE OF INVENTION: AND TREATMENT OF B. MICROTI INFECTION
; FILE REFERENCE: 210121.42604

GenCore version 5.1.3
Copyright (c) 1993 - 2002 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: November 27, 2002, 22:28:40 ; Search time 5666 seconds
(without alignments)
11649.341 Million cell updates/sec

Title: US-09-642-034-4
Perfect score: 2268
Sequence: 1 atggcaggaggttatctgt.....tgtttcgtataaatttctag 2268

Scoring table: IDENTITY_NUC

Gapop 10.0 ; Gapext 1.0

Searched: 2054640 seqs, 14551402878 residues

Total number of hits satisfying chosen parameters: 4109280

Minimum DB seq length: 0
Maximum DB seq length: 20000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : GenBank :

- 1: gb_ba.*
- 2: gb_htg.*
- 3: gb_in.*
- 4: gb_ov.*
- 5: gb_ov.*
- 6: gb_pat.*
- 7: gb_pl.*
- 8: gb_pl.*
- 9: gb_pr.*
- 10: gb_ro.*
- 11: gb_sts.*
- 12: gb_sy.*
- 13: gb_un.*
- 14: gb_vi.*
- 15: em_ba.*
- 16: em_fun.*
- 17: em_hum.*
- 18: em_in.*
- 19: em_mu.*
- 20: em_mu.*
- 21: em_or.*
- 22: em_ov.*
- 23: em_pat.*
- 24: em_ph.*
- 25: em_pl.*
- 26: em_ro.*
- 27: em_sts.*
- 28: em_un.*
- 29: em_vi.*
- 30: em_htg_hum.*
- 31: em_htg_inv.*
- 32: em_htg_other.*
- 33: em_htg_mus.*
- 34: em_htg_pln.*
- 35: em_htg_rod.*
- 36: em_htg_mam.*
- 37: em_htg_vrt.*
- 38: em_sy.*
- 39: em_htgo_hum.*
- 40: em_htgo_mus.*
- 41: em_htgo_other.*

Pred. No. is the number of results predicted by chance to have a

score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Match %	Length	DB	ID	Description
1	2268	100.0	3461	6	AX465588	Sequence
2	2264.8	99.9	2776	6	AX207207	Sequence
3	2222	98.0	2744	9	HSU41060	Homo sapien
4	2198	96.9	3461	6	AX207205	Sequence
5	1616	71.3	3287	10	AB071697	Sequence
6	1263	55.7	1310	6	AB071691	Mus muscu
7	1263	55.7	2404	6	I76892	Sequence 1
8	1151	50.7	2236	9	BC008317	Sequence 3
9	789	34.8	160170	2	AP001158	Homo sapi
10	787.4	34.7	221941	9	AC091060	Homo sapi
11	787.4	34.7	224788	2	AP001905	Homo sapi
12	619.6	27.3	899	6	AX465590	Sequence
13	549.4	22.5	551	6	AX207216	Sequence
14	509.4	22.5	1597	6	AX017261	Sequence
15	421.4	18.6	193437	2	AC126667	Sequence
16	319.2	14.1	337	6	AX393861	Rattus no
17	319	14.1	5231	9	AB033091	Sequence
18	319	14.1	5322	9	AB033091	Homo sapi
19	273.4	12.1	61103	2	AC131276	Sequence
20	191.2	8.4	193437	2	AC126667	Homo sapi
21	144	6.3	52684	2	AC014796	Rattus no
22	144	6.3	174712	3	AC010041	Drosophi
23	144	6.3	195148	3	AC010019	Drosophi
24	144	6.3	282267	3	AE003555	Drosophi
25	129.2	5.7	2174	10	BC021530	Mus muscu
26	128.4	5.7	2015	9	AB051127	Macaca fa
27	127.6	5.6	4573	6	AX409002	Sequence
28	127.6	5.6	4573	9	HUMOREKGLP	Human mRNA
29	126	5.6	4610	6	AX078294	Sequence
30	126	5.6	4698	6	AX281784	Sequence
31	123.4	5.4	1383	6	AX118897	Sequence
32	123.4	5.4	1803	9	AF193052	Homo sapi
33	123.4	5.4	2604	9	AK027652	Homo sapi
34	123.4	5.4	3169	9	BC012125	Homo sapi
35	123.4	5.4	3246	9	AB040120	Homo sapi
36	122.6	5.4	1895	9	BC027884	Homo sapi
37	119.2	5.3	1755	9	BC015770	Homo sapi
38	119.2	5.3	55577	2	AC014109	Drosophi
39	119.2	5.3	170071	3	AC093497	Drosophi
40	119.2	5.3	194979	3	AC091209	Drosophi
41	119.2	5.3	295377	3	AE003532	Drosophi
42	118	5.2	150	9	HUMDSF04M5	Human HepG2
43	116.2	5.1	1931	10	BC028990	Mus muscu
44	115.8	5.1	3060	10	BC006731	Mus muscu
45	110.6	4.9	1888	6	AX061633	Sequence

ALIGNMENTS

RESULT 1
AX465588
LOCUS AX465588
DEFINITION Sequence 1 from Patent WO0216939.
ACCESSION AX465588
VERSION AX465588.1 GI:21899891
KEYWORDS human.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
Mammalia; Euthera; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1
AUTHORS Mack D., Gish, K.C. and Wilson, K.E.
TITLE Methods of diagnosis of cancer and screening for cancer modulators
JOURNAL Patent: WO 0216939-A 1 28-FEB-2002;

AX465588 3461 bp DNA linear PAT 16-JUL-2002

DB	1818	TACCATCATATTTCTCCATCATCACCAGTACCAAAACACCATCCCTCACATACACAGCAG	1877
QY	1741	CGCTACTCTCGGAGGAGCTGAAAGATTCGGCGCTGCCACTTTGGCCCTGGATGTGCATA	1800
DB	1878	CGCTACTCTCGGAGGAGCTGAAAGATTCGGCGCTGCCACTTTGGCCCTGGATGTGCATA	1937
QY	1801	ATGGGTGATGAGCTTACACAAATTCACCGATGCCCTAGCAATTTGGTGTCTGTTTATCTAA	1860
DB	1938	ATGGGTGATGAGCTTACACAAATTCACCGATGCCCTAGCAATTTGGTGTCTGTTTATCTAA	1997
QY	1861	GGCTTATCAAGTGGTAACTTCTGTTGCTGTCTGCTGCATCAGTTGCTGTCTATCAA	1920
DB	1998	GGCTTATCAAGTGGTAACTTCTGTTGCTGTCTGCTGCATCAGTTGCTGTCTATCAA	2057
QY	1921	TTTATGTGACCTTTGCTGTCTTACTTAAAGGTGGCATGACCGTTTAAAGCATGCTGCTTTAT	1980
DB	2058	TTTATGTGACCTTTGCTGTCTTACTTAAAGGTGGCATGACCGTTTAAAGCATGCTGCTTTAT	2117
QY	1981	AATGCATTTCTACGCCATGCTGGCTATCTTGAATGGCAACAGAGAAATTTTCATGCTCAT	2040
DB	2118	AATGCATTTCTACGCCATGCTGGCTATCTTGAATGGCAACAGAGAAATTTTCATGCTCAT	2177
QY	2041	TATGCTGAAATCTTTCTATGCTGGATATTTGCACCTTACTGCTGGCTTATTCATGATGTT	2100
DB	2178	TATGCTGAAATCTTTCTATGCTGGATATTTGCACCTTACTGCTGGCTTATTCATGATGTT	2237
QY	2101	CTCTCTGCTTATGATGCTTACCTGAAATGCTGCACAAATGATGCTGATGACATGATGAC	2160
DB	2238	CTCTCTGCTTATGATGCTTACCTGAAATGCTGCACAAATGATGCTGATGACATGATGAC	2297
QY	2161	CGCTCGGCGTATTTCTTTTACAGAAATGCTGGGATGCTTTTGGGTTTTCGAAATTTATGTTA	2220
DB	2298	CGCTCGGCGTATTTCTTTTACAGAAATGCTGGGATGCTTTTGGGTTTTCGAAATTTATGTTA	2357
QY	2221	CTTATTTCCATATTTGAACATAAAATCGTGTGTTTCGTATATAAATTTCTAG	2268
DB	2358	CTTATTTCCATATTTGAACATAAAATCGTGTGTTTCGTATATAAATTTCTAG	2405
RESULT 2			
AX207207		2776 bp	DNA
SEQUENCE	3 from Patent WO0155178		linear
DEFINITION	AX207207		
ACCESSION	AX207207.1	GI:15394961	
VERSION			
KEYWORDS			
SOURCE			
ORGANISM			
REFERENCE			
AUTHORS			
TITLE			
JOURNAL			
FEATURES			
source			
BASE COUNT	823 a	583 c	592 g
ORIGIN			
Query Match			
Best Local Similarity	99.9%	Score 2264.8	DB 6
Matches 2266	Conservative	0	Mismatches
		0	Indels
		0	Gaps
		0	
QY	1	ATGCGGAGGAAGTTATCTGTAATCTTGATGCTGACCTTTTGGCCCTCTCTGTGCACAAATCCC	60
DB	73	ATGCGGAGGAAGTTATCTGTAATCTTGATGCTGACCTTTTGGCCCTCTCTGTGCACAAATCCC	132
QY	61	CTTCATCAACTAAAGCAGCTGTGTTCCCGCAGACCACTGAGAAATTTAGTCCGAATGG	120

Db 318 TTTCTACCGTATCGGAAAATAATTTCTTTGTTCAGTTGAGGGTTTCAGAAAAAT1ACTTCAA 377
|||||
QY 241 AATATAGGCATAGATAAGATTAAAGAAATCCATATACACCATGACGACGACCATCATCTCA 300
|||||
Db 378 AATATAGGCATAGATAAGATTAAAGAAATCCATATACACCATGACGACGACCATCATCTCA 437
|||||
QY 301 GATCAGTATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 360
|||||
Db 418 GATCAGTATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 497
|||||
QY 361 GATCAGTATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 420
|||||
Db 498 GATCAGTATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 539
|||||
QY 421 AAGCGAAAAGCTCTTTTGGCCGACCATGACTCAGATAGTCTCAGGTAAAGATCTTCAAGAAC 480
|||||
Db 540 AAGCGAAAAGCTCTTTTGGCCGACCATGACTCAGATAGTCTCAGGTAAAGATCTTCAAGAAC 599
|||||
QY 481 AGCGAGGGAAGAGAGTCTACCGGACGAGAACATGCGGTAGAGGAATGTCAAGGAC 540
|||||
Db 600 AGCGAGGGAAGAGAGTCTACCGGACGAGAACATGCGGTAGAGGAATGTCAAGGAC 659
|||||
QY 541 ACTGTTAGTGTCTAGTGAAGTGAACCTCAACTGTGTACACACTGTCTCTCAAGGAATCTCAC 600
|||||
Db 660 AGTGTAGTGTCTAGTGAAGTGAACCTCAACTGTGTACACACTGTCTCTCAAGGAATCTCAC 719
|||||
QY 601 TTTCTAGAGACAATAGAGACTCCAAAGACTCGGAAACTCTTTCCGCAAAAGATGTAAAGAGC 660
|||||
Db 720 TTTCTAGAGACAATAGAGACTCCAAAGACTCGGAAACTCTTTCCGCAAAAGATGTAAAGAGC 779
|||||
QY 661 TCCACTCCACCCAGTGTACATCAAGAGCGCGGTGAGCGCGGTGCGGTGAGGAAACA 720
|||||
Db 780 TCCACTCCACCCAGTGTACATCAAGAGCGCGGTGAGCGCGGTGCGGTGAGGAAACA 839
|||||
QY 721 AATGAATCTGTAGTGAAGTGAACCTCAACTGTGTACACACTGTCTCTCAAGGAATCTCAC 780
|||||
Db 840 AATGAATCTGTAGTGAAGTGAACCTCAACTGTGTACACACTGTCTCTCAAGGAATCTCAC 899
|||||
QY 781 CTTAGGAGTGTCTCAATGCAATGCAATGCAATGCAATGCAATGCAATGCAATGCAATGCAATGCA 840
|||||
Db 900 CTTAGGAGTGTCTCAATGCAATGCAATGCAATGCAATGCAATGCAATGCAATGCAATGCAATGCA 959
|||||
QY 841 CCGCTCAATGCAATGCAATGCAATGCAATGCAATGCAATGCAATGCAATGCAATGCAATGCAATGCA 900
|||||
Db 960 CCGCTCAATGCAATGCAATGCAATGCAATGCAATGCAATGCAATGCAATGCAATGCAATGCAATGCA 1019
|||||
QY 901 AGATCTTCTCTGATTCATCAAGTGAAGAGGCTGAAATCCCTCCAAAGACCTATTCA 960
|||||
Db 1020 AGATCTTCTCTGATTCATCAAGTGAAGAGGCTGAAATCCCTCCAAAGACCTATTCA 1079
|||||
QY 961 TTACAAATAGCTCGGTTGGTGGTTTATAGCCATTTTCCATCATCATGATGATGATGATGATGATGATG 1020
|||||
Db 1080 TTACAAATAGCTCGGTTGGTGGTTTATAGCCATTTTCCATCATCATGATGATGATGATGATGATGATG 1139
|||||
QY 1021 CTGGGGGTATCTTAGTGCCTCTCATGAATCGGGTGTCTTCAAAATCTCTCTGAGTTTC 1080
|||||
Db 1140 CTGGGGGTATCTTAGTGCCTCTCATGAATCGGGTGTCTTCAAAATCTCTCTGAGTTTC 1199
|||||
QY 1081 TTTTGGGCACTGGCGTTGGGCACTTTTGGGCACTTTTGGGCACTTTTGGGCACTTTTGGGCACTTTT 1140
|||||
Db 1200 TTTTGGGCACTGGCGTTGGGCACTTTTGGGCACTTTTGGGCACTTTTGGGCACTTTTGGGCACTTTT 1259
|||||
QY 1141 TCTCATGCAAGTCAACCATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 1200
|||||
Db 1260 TCTCATGCAAGTCAACCATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 1319
|||||
QY 1201 GGACCATTTTTCAGTCACTGCTCTTCTCAAAACATGCAAAACTGCTATTTTGGTATTC 1260
|||||
Db 1320 GGACCATTTTTCAGTCACTGCTCTTCTCAAAACATGCAAAACTGCTATTTTGGTATTC 1379
|||||
QY 1261 ACCTGGAGGCTCTAACAGCTCTAGGAGGCTGTATTTTCATGTTCTTGTGTAACATGTC 1320
|||||

Db 1380 ACGTGAAGGGTCTAACAGCTCTAGAGGGCTGTATTTTCATGTTTCTTGTGTAACATGTC 1439
|||||
QY 1321 CTCACATTTGATCAAAACAATTTAAAGATTAAGAGAAAAGAAATCAGAAGAAACCTGAAAAT 1380
|||||
Db 1440 CTCACATTTGATCAAAACAATTTAAAGATTAAGAGAAAAGAAATCAGAAGAAACCTGAAAAT 1499
|||||
QY 1381 GAT 1440
|||||
Db 1500 GAT 1559
|||||
QY 1441 GAGGAGAAAGTATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 1500
|||||
Db 1560 GAGGAGAAAGTATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 1619
|||||
QY 1501 CCCTCCCACTTTGATTTCTCAGCAGCCTGCGAGTCTTGGAGAAAGAGGTCATGATAGCT 1560
|||||
Db 1620 CCCTCCCACTTTGATTTCTCAGCAGCCTGCGAGTCTTGGAGAAAGAGGTCATGATAGCT 1679
|||||
QY 1561 CATGCTCATCCACAGGAGTCTACAATGAATATGTACCCAGAGGGTGCAAGAATAAATGC 1620
|||||
Db 1680 CATGCTCATCCACAGGAGTCTACAATGAATATGTACCCAGAGGGTGCAAGAATAAATGC 1739
|||||
QY 1621 CATTCACATTTCCACGATACACTCGGCCAGTCAGACGATCTCATTCACCACTCATGAC 1680
|||||
Db 1740 CATTCACATTTCCACGATACACTCGGCCAGTCAGACGATCTCATTCACCACTCATGAC 1799
|||||
QY 1681 TACCATCATATTTCCATCATCATCACCACCAACCAACCACTCATTCACCACTCATGAC 1740
|||||
Db 1800 TACCATCATATTTCCATCATCATCACCACCAACCAACCACTCATTCACCACTCATGAC 1859
|||||
QY 1741 CGCTACTCTCGGAGGAGCTGAAAGATGCGCGGCTGCGCACTTTGGCCTCGATGGTGATA 1800
|||||
Db 1860 CGCTACTCTCGGAGGAGCTGAAAGATGCGCGGCTGCGCACTTTGGCCTCGATGGTGATA 1919
|||||
QY 1801 ATGGGTGATGGCTGCAAAATTTTACGGATGGCTTACCAATTTGGTGGTCTTTACTGAA 1860
|||||
Db 1920 ATGGGTGATGGCTGCAAAATTTTACGGATGGCTTACCAATTTGGTGGTCTTTACTGAA 1979
|||||
QY 1861 GGCTTATCAAGTGGTTTAAAGTACTTCTGTTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1920
|||||
Db 1980 GGCTTATCAAGTGGTTTAAAGTACTTCTGTTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 2039
|||||
QY 1921 TTAGGTGACTTTGCTGTTTCTTAAAGGCTGGCATGACCGTTAAGCAGGCTGCTCTTTAT 1980
|||||
Db 2040 TTAGGTGACTTTGCTGTTTCTTAAAGGCTGGCATGACCGTTAAGCAGGCTGCTCTTTAT 2099
|||||
QY 1981 AATGCAATTTGTCGCCATGCTGGGTATCTTGGAAATGGCAACAGCAATTTTTCATTTGGTCAT 2040
|||||
Db 2100 AATGCAATTTGTCGCCATGCTGGGTATCTTGGAAATGGCAACAGCAATTTTTCATTTGGTCAT 2159
|||||
QY 2041 TATGCTGAAAATGTTTCTTATGTTGATATTTGACATTTTACTGCTGCTGCTGCTGCTGCTGCTGCT 2100
|||||
Db 2160 TATGCTGAAAATGTTTCTTATGTTGATATTTGACATTTTACTGCTGCTGCTGCTGCTGCTGCTGCT 2219
|||||
QY 2101 GCTCTGGTGTATGTTGATGTTGATGTTGATGTTGATGTTGATGTTGATGTTGATGTTGATGTTGAT 2160
|||||
Db 2220 GCTCTGGTGTATGTTGATGTTGATGTTGATGTTGATGTTGATGTTGATGTTGATGTTGATGTTGAT 2279
|||||
QY 2161 CGCTGGGGGTATTTCTTTTACAGAAATGCTGGGATGCTTTTGGGTTTGGGTTTGGGTTTGGGTTT 2220
|||||
Db 2280 CGCTGGGGGTATTTCTTTTACAGAAATGCTGGGATGCTTTTGGGTTTGGGTTTGGGTTTGGGTTT 2339
|||||
QY 2221 CTTTATTTCCATATTTTGAACATAAAATCGTGTTCGTGATATAAATTTCTAG 2268
|||||
Db 2340 CTTTATTTCCATATTTTGAACATAAAATCGTGTTCGTGATATAAATTTCTAG 2387
|||||

RESULT 4
AX207205
LOCUS
DEFINITION
ACCESSION
VERSION
AX207205
Sequence 1 from Patent WO0155178.
AX207205
AX207205.1 GI:15394960
3461 bp
DNA
Linear
PAT 30-AUG-2001

Quality coverage: 10.41x in Q20 bases; sum-of-contigs

NOTE: This is a 'working draft' sequence. It currently consists of 15 contigs. The true order of the pieces is not known and their order in this sequence record is arbitrary. Gaps between the contigs are represented as runs N, but the exact sizes of the gaps are unknown. This record will be updated with the finished sequence as soon as it is available and the accession number will be preserved.

	1	2590	contig	of	2590	bp	in	length
	29691	43906	contig	of	20216	bp	in	length
	50007	69120	contig	of	19114	bp	in	length
	65221	84192	contig	of	14972	bp	in	length
	84293	99890	contig	of	15598	bp	in	length
	99991	110467	contig	of	10477	bp	in	length
	110568	118491	contig	of	7924	bp	in	length
	118592	128449	contig	of	7858	bp	in	length
	126550	134958	contig	of	5995	bp	in	length
	135059	141053	contig	of	5995	bp	in	length
	141154	145920	contig	of	4767	bp	in	length
	146021	149439	contig	of	3419	bp	in	length
	149540	154173	contig	of	4634	bp	in	length
	154274	157616	contig	of	3343	bp	in	length
	157717	160170	contig	of	2454	bp	in	length

* NOTE: This is a 'working draft' sequence. It currently
* consists of 15 contigs. The true order of the pieces is
* not known, and their order in this sequence is
* arbitrary. Gaps between the contigs are represented as
* runs of 'N', but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence.
* As soon as it is available and the accession number will
* be preserved.

*	1	29590:	contig of 29590 bp in length
*	25991	29690:	gap of 100 bp
*	23691	49906:	contig of 20216 bp in length
*	49007	50006:	gap of 100 bp
*	50007	69120:	contig of 19114 bp in length
*	69121	69220:	gap of 100 bp
*	69221	84192:	contig of 14972 bp in length
*	84193	84292:	gap of 100 bp
*	84293	99890:	contig of 15598 bp in length
*	99891	99990:	gap of 100 bp
*	99991	110467:	contig of 10477 bp in length
*	110468	110567:	gap of 100 bp
*	110568	118491:	contig of 7924 bp in length
*	118492	118591:	gap of 100 bp
*	118592	1264449:	contig of 7858 bp in length
*	126450	126549:	gap of 100 bp
*	126550	134958:	contig of 8409 bp in length
*	134959	135058:	gap of 100 bp
*	135059	140553:	contig of 5995 bp in length
*	141054	141153:	gap of 100 bp
*	141154	145920:	contig of 4767 bp in length
*	145921	146020:	gap of 100 bp
*	146021	149439:	contig of 3419 bp in length
*	149440	149539:	gap of 100 bp
*	149540	154173:	contig of 4634 bp in length
*	154174	154273:	gap of 100 bp
*	154274	157616:	contig of 3343 bp in length
*	157617	157716:	gap of 100 bp
*	157717	160170:	contig of 2454 bp in length.

FEATURES	SOURCE
1. The first two columns are labeled "FEATURES" and "SOURCE".	
2. The third column is labeled "FEATURES".	
3. The fourth column is labeled "SOURCE".	
4. The fifth column is labeled "FEATURES".	
5. The sixth column is labeled "SOURCE".	
6. The seventh column is labeled "FEATURES".	
7. The eighth column is labeled "SOURCE".	
8. The ninth column is labeled "FEATURES".	
9. The tenth column is labeled "SOURCE".	
10. The eleventh column is labeled "FEATURES".	
11. The twelfth column is labeled "SOURCE".	
12. The thirteenth column is labeled "FEATURES".	
13. The fourteenth column is labeled "SOURCE".	
14. The fifteenth column is labeled "FEATURES".	
15. The sixteenth column is labeled "SOURCE".	
16. The seventeenth column is labeled "FEATURES".	
17. The eighteenth column is labeled "SOURCE".	
18. The nineteenth column is labeled "FEATURES".	
19. The twentieth column is labeled "SOURCE".	
20. The twenty-first column is labeled "FEATURES".	
21. The twenty-second column is labeled "SOURCE".	
22. The twenty-third column is labeled "FEATURES".	
23. The twenty-fourth column is labeled "SOURCE".	
24. The twenty-fifth column is labeled "FEATURES".	
25. The twenty-sixth column is labeled "SOURCE".	
26. The twenty-seventh column is labeled "FEATURES".	
27. The twenty-eighth column is labeled "SOURCE".	
28. The twenty-ninth column is labeled "FEATURES".	
29. The thirtieth column is labeled "SOURCE".	
30. The thirty-first column is labeled "FEATURES".	
31. The thirty-second column is labeled "SOURCE".	
32. The thirty-third column is labeled "FEATURES".	
33. The thirty-fourth column is labeled "SOURCE".	
34. The thirty-fifth column is labeled "FEATURES".	
35. The thirty-sixth column is labeled "SOURCE".	
36. The thirty-seventh column is labeled "FEATURES".	
37. The thirty-eighth column is labeled "SOURCE".	
38. The thirty-ninth column is labeled "FEATURES".	
39. The fortieth column is labeled "SOURCE".	
40. The forty-first column is labeled "FEATURES".	
41. The forty-second column is labeled "SOURCE".	
42. The forty-third column is labeled "FEATURES".	
43. The forty-fourth column is labeled "SOURCE".	
44. The forty-fifth column is labeled "FEATURES".	
45. The forty-sixth column is labeled "SOURCE".	
46. The forty-seventh column is labeled "FEATURES".	
47. The forty-eighth column is labeled "SOURCE".	
48. The forty-ninth column is labeled "FEATURES".	
49. The fiftieth column is labeled "SOURCE".	
50. The fifty-first column is labeled "FEATURES".	
51. The fifty-second column is labeled "SOURCE".	
52. The fifty-third column is labeled "FEATURES".	
53. The fifty-fourth column is labeled "SOURCE".	
54. The fifty-fifth column is labeled "FEATURES".	
55. The fifty-sixth column is labeled "SOURCE".	
56. The fifty-seventh column is labeled "FEATURES".	
57. The fifty-eighth column is labeled "SOURCE".	
58. The fifty-ninth column is labeled "FEATURES".	
59. The sixtieth column is labeled "SOURCE".	
60. The sixty-first column is labeled "FEATURES".	
61. The sixty-second column is labeled "SOURCE".	
62. The sixty-third column is labeled "FEATURES".	
63. The sixty-fourth column is labeled "SOURCE".	
64. The sixty-fifth column is labeled "FEATURES".	
65. The sixty-sixth column is labeled "SOURCE".	
66. The sixty-seventh column is labeled "FEATURES".	
67. The sixty-eighth column is labeled "SOURCE".	
68. The sixty-ninth column is labeled "FEATURES".	
69. The seventieth column is labeled "SOURCE".	
70. The seventy-first column is labeled "FEATURES".	
71. The seventy-second column is labeled "SOURCE".	
72. The seventy-third column is labeled "FEATURES".	
73. The seventy-fourth column is labeled "SOURCE".	
74. The seventy-fifth column is labeled "FEATURES".	
75. The seventy-sixth column is labeled "SOURCE".	
76. The seventy-seventh column is labeled "FEATURES".	
77. The seventy-eighth column is labeled "SOURCE".	
78. The seventy-ninth column is labeled "FEATURES".	
79. The eightieth column is labeled "SOURCE".	
80. The eighty-first column is labeled "FEATURES".	
81. The eighty-second column is labeled "SOURCE".	
82. The eighty-third column is labeled "FEATURES".	
83. The eighty-fourth column is labeled "SOURCE".	
84. The eighty-fifth column is labeled "FEATURES".	
85. The eighty-sixth column is labeled "SOURCE".	
86. The eighty-seventh column is labeled "FEATURES".	
87. The eighty-eighth column is labeled "SOURCE".	
88. The eighty-ninth column is labeled "FEATURES".	
89. The ninetieth column is labeled "SOURCE".	
90. The ninety-first column is labeled "FEATURES".	
91. The ninety-second column is labeled "SOURCE".	
92. The ninety-third column is labeled "FEATURES".	
93. The ninety-fourth column is labeled "SOURCE".	
94. The ninety-fifth column is labeled "FEATURES".	
95. The ninety-sixth column is labeled "SOURCE".	
96. The ninety-seventh column is labeled "FEATURES".	
97. The ninety-eighth column is labeled "SOURCE".	
98. The ninety-ninth column is labeled "FEATURES".	
99. The hundredth column is labeled "SOURCE".	

```
misc_feature      1. .29590
                  /note="assembly_fragment"
misc_feature      29691. .49906
                  /note="assembly_fragment"
misc_feature      50007. .69120
                  /note="assembly_fragment"
```

QY	1998	TCTGTCAATGAGTTCATGAAATTTAGCTTGCTTCTACTAAAGCGCTGGCATGA	1957
Db	781	TT	
Db	781	TCTGTCAATGAGTTCATGAAATTTAGCTTGCTTCTACTAAAGCGCTGGCATGA	840
QY	1958	CCCTTTAAGCAGCGCTGCCCTTTAATATGCAATTCAGGCATGCTCAGGCATGCTTGTGGAATGG	2017
Db	841	TT	
Db	841	CCCTTTAAGCAGCGCTGCCCTTTAATATGCAATTCAGGCATGCTCAGGCATGCTTGTGGAATGG	900
QY	2018	CAACAGGAAATTTTCATTTGCTCATTTATGCTGAAAATGTTTCTATCTGCATATTTGCACTTA	2077
Db	901	CAACAGCAATTTTCATTTGCTCATTTATGCTGAAAATGTTTCTATCTGCATATTTGCACTTA	960
QY	2078	CTGCTGGCTTTATTCATGATATGTTGCTCTGGTGTGATATGGTACCTGAAATGCTGCACAAATG	2137
Db	961	TT	
Db	961	CTGCTGGCTTTATTCATGATATGTTGCTCTGGTGTGATATGGTACCTGAAATGCTGCACAAATG	1020
QY	2138	ATGCTAGTCACCATGGATGCTAGCCGCTGGGGGTATTTCTTTTTACAGAAATGCTTGGCATGC	2197
Db	1021	TT	
Db	1021	ATGCTAGTCACCATGGATGCTAGCCGCTGGGGGTATTTCTTTTTACAGAAATGCTTGGCATGC	1080
QY	2198	TTTTTGGGTTTTTGGAAATATGTTACTTTATTTCCATTTTGAACATAAAATGCTTTTCGTA	2257
Db	1081	TTTTTGGGTTTTTGGAAATATGTTACTTTATTTCCATTTTGAACATAAAATGCTTTTCGTA	1140
QY	2258	TAAATTTCTAG	2268
Db	1141	TT	
Db	1141	TAAATTTCTAG	1151

[illegible]

REFERENCE AUTHORS	Mammalia: Euthera: Primates: Catarrhini: Homiidae: Homo. 1 (bases 1 to 160170)
TITLE	Hattori,M., Ishii,K., Toyoda,A., Taylor,T.D., Hong-Seoq,P., Fujiyama,A., Yada,T., Totoki,Y., Watanabe,H. and Sakaki,Y.
JOURNAL	Homo sapiens 160170 genomic DNA of 18q12
REFERENCE	Published Only in DataBase (2000)
REFERENCE AUTHORS	2 (bases 1 to 160170)
TITLE	Hattori,M., Ishii,K., Toyoda,A., Taylor,T.D., Hong-Seoq,P., Fujiyama,A., Yada,T., Totoki,Y., Watanabe,H. and Sakaki,Y.
JOURNAL	Direct Submission Submitted (10-FEB-2000) Masahira Hattori, The Institute of Physical and Chemical Research (RIKEN), Genomic Sciences Center (GSC); 1-15-1 Kitasato, Saitamahara, Kanagawa 228-8555, Japan (E-mail:hattori@gsc.riken.go.jp, Url:http://hgp.qst.riken.go.jp/ Tel:81-42-778-9923, Fax:81-42-778-9924) On Jul 14, 2000 this sequence version replaced qi:8118726.

```

Center: RIKEN Genomic Sciences Center(GSC)
Center code: RIKEN
Web site: http://hgp.gsc.riken.go.jp/
Contact: hatfori@gs.riken.go.jp
----- Project Information -----
Center project name: HumDraft18
Center clone name: RP11-701c9
----- Summary Statistics -----
Sequencing vector: PCR products; 100% of reads
Chemistry: Dye-terminator ET-amersham; 100% of reads
Assembly program: Phrap; version 0.950329
Consensus quality: 155820 bases at least Q40
Consensus quality: 157459 bases at least Q30
Consensus quality: 158091 bases at least Q20
Insert size: 158770; sum-of-contigs

```


Koman, J., Rosetti, M., Roy, A., Santos, R., Schaner, S., Schupback, R., Seaman, S., Severy, P., Spencer, B., Stange-Thomann, N., Stojanovic, N., Strauss, N., Subramanian, A., Talamas, J., Tesfaye, S., Theodore, J., Topliam, K., Travers, M., Travis, N., Trigilio, J., Vassiliev, H., Viel, R., Vo, A., Wilson, B., Wu, X., Wyman, D., Ye, W.-J., Young, G., Zainoun, J., Zembek, L., Zimmer, A. and Zody, M.

Direct Submission
Submitted (28-NOV-2001) Whitehead Institute/MIT Center for Genome Research, 320 Charles Street, Cambridge, MA 02141, USA
On Nov 4, 2001 this sequence version replaced gi:16118168.
All repeats were identified using RepeatMasker:
Smit, A.F.A. & Green, P. (1996-1997)
<http://ftp.genome.washington.edu/RM/RepeatMasker.html>

----- Genome Center
Center: Whitehead Institute/ MIT Center for Genome Research

Center code: WIBR
Web site: <http://www-seq.wi.mit.edu>
Contact: sequence_submissions@genome.wi.mit.edu
----- Project Information
Center project name: L12269
Center clone name: 723_J_4

FEATURES

source

location/Qualifiers
1. .221941
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="18"
/map="18"
/clone="RP11-723J4"
/clone_lib="RPC1-11 Human Male BAC"

240. .261
/note="x30 qual SINGL region"

322. .326
/note="x30 qual SINGL region"

635. .670
/rpt_family="L1ME2"
complement(674. .963)

688. .715
/rpt_family="AluSg"

951. .955
/note="x30 qual SINGL region"

1010. .1013
/note="x30 qual SINGL region"

1124. .1441
/rpt_family="AluSx"

2833. .2926
/rpt_family="AluSx"

complement(2960. .3257)

complement(3291. .3371)

3807. .4209
/rpt_family="L1MD1"

complement(4298. .4597)

5434. .5597
/rpt_family="AluSx"

complement(6415. .6580)

complement(7324. .7451)

7505. .7750
/rpt_family="MLT1H"

complement(7768. .7984)

8022. .8284
/rpt_family="MLT1J"

complement(8865. .9153)

9584. .9680
/rpt_family="AluSc"

complement(9680. .9680)

11358. .11524
/rpt_family="AluSg/x"

complement(11524. .11524)

complement(11524. .11524)

complement(11524. .11524)

complement(11524. .11524)

complement(11524. .11524)

complement(11524. .11524)

complement(11524. .11524)

complement(11524. .11524)

complement(11524. .11524)

complement(11524. .11524)

complement(11524. .11524)

complement(11524. .11524)

complement(11524. .11524)

complement(11524. .11524)

complement(11524. .11524)

complement(11524. .11524)

complement(11524. .11524)

complement(11524. .11524)

complement(11524. .11524)

complement(11524. .11524)

complement(11524. .11524)

complement(11524. .11524)

complement(11524. .11524)

/rpt_family="AluSx"
11555. .11586
/rpt_family="(TAAA)n"
11587. .11750
/rpt_family="L1ME2"
11838. .12141
/rpt_family="AluSx"
complement(12195. .12762)
/rpt_family="MLT1E2"
12824. .12950
/rpt_family="AluSg/x"
12955. .13254
/rpt_family="AluSx"
13256. .13920
/rpt_family="L1MB2"
13921. .14181
/rpt_family="AluSx"
14182. .14516
/rpt_family="L1MB2"
14572. .14886
/rpt_family="L1ME2"
14905. .14936
/rpt_family="MERS8B"
14937. .15226
/rpt_family="AluSg"
15227. .15376
/rpt_family="MERS8B"
15377. .16032
/rpt_family="L1ME2"
complement(16039. .16172)
/rpt_family="L1MCA"
complement(16197. .17024)
/rpt_family="L1MCA"
complement(17025. .17305)
/rpt_family="AluSx"
complement(17306. .17710)
/rpt_family="L1MCA"
17710. .18046
/rpt_family="L1ME2"
18047. .18330
/rpt_family="GA-rich"
18336. .19013
/rpt_family="L1ME2"
19014. .19286
/rpt_family="AluYa5"
19287. .19895
/rpt_family="L1ME2"
19896. .20066
/rpt_family="AluJb"
20070. .20562
/rpt_family="L1ME2"
20580. .20722
/rpt_family="L1ME2"
20817. .20961
/rpt_family="L1MCS"
complement(20990. .21142)
/rpt_family="L1MCS"
21143. .21430
/rpt_family="AluJo"
complement(21431. .21443)
/rpt_family="L1MCS"
21444. .21466
/rpt_family="(TAAA)n"
complement(21467. .21637)
/rpt_family="L1MCS"
complement(21634. .21943)
/rpt_family="AluJb"
complement(22021. .22069)
/rpt_family="L1MA8"
22070. .22092
/rpt_family="(TA)n"
complement(22093. .22227)

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

Query Match	34.7%	Score 787.4	DB 9	Length 221941
Rest Local Similarity	99.9%	Pred. No. 3.7e-184		
Matches 788	Conservative	Mismatches 1	Indels 0	Gaps 0
QY 1	ATGCGCAGCAAGTATCTGTAATCTTGATCCTCAGCCTTTGGCCCTCTCTCTGCAACAATCCC	60		
Db 187366	ATGCGGAGGAGTTATCTGTAATCTTGATCCTCAGCCTTTGGCCCTCTCTCTGCAACAATCCC	187307		
QY 61	CTTCATGACATTAAGCAGCTGCTTTCCGCCAGACCTTCAGAGAAATTAAGTCCGAATTCGG	120		
Db 187306	CTTCATGACATTAAGCAGCTGCTTTCCGCCAGACCTTCAGAGAAATTAAGTCCGAATTCGG	187247		
QY 121	GAATCTGCATTAATCTTGACCTTGCGAATTTCCACAGCGCAATATCATCTCAAGACATTT	180		
Db 187246	GAATCTGCATTAATCTTGACCTTGCGAATTTCCACAGCGCAATATCATCTCAAGACATTT	187187		
QY 181	TTCTACCGCTATGAGAGAAATAATCTTTGTCTAGTTGAAGGGTTGAGAAAATTTACTTCAA	240		
Db 187186	TTCTACCGCTATGAGAGAAATAATCTTTGTCTAGTTGAAGGGTTGAGAAAATTTACTTCAA	187127		
QY 241	AATATAGGCATAGATAGATTAAAGAAATCCATATACCATCAGCAGCATCATCTCA	300		
Db 187126	AATATAGGCATAGATAGATTAAAGAAATCCATATACCATCAGCAGCATCATCTCA	187067		
QY 401	GACCAACAGCATCATCAGACCATGAGCGTCACTCAGACCATGAGCATCATCTCA	360		
Db 187066	GACCAACAGCATCATCAGACCATGAGCGTCACTCAGACCATGAGCATCATCTCA	187007		
QY 361	GAGCATCTCTCTGACCATGATCATCACTCTCAGCATAATCATGCTCTCTGCTTAAAT	420		
Db 187006	GAGCATCTCTCTGACCATGATCATCACTCTCAGCATAATCATGCTCTCTGCTTAAAT	186947		
QY 421	AACGGAAGATCTTTTGGCCAGCACCATGACTCAGATGATTCAGGTAAGATCTTAGAAC	480		
Db 186946	AACGGAAGATCTTTTGGCCAGCACCATGACTCAGATGATTCAGGTAAGATCTTAGAAC	186887		
QY 481	AGCAGGGAAGAGAGCTCAGCAGCAGCAACATGCCAGTGGTAGAAGAAATCTCAAGAC	540		
Db 186886	AGCAGGGAAGAGAGCTCAGCAGCAGCAACATGCCAGTGGTAGAAGAAATCTCAAGAC	186827		
QY 541	AGTGTATAGTCTAGTGAAGTGACCTCAACTGTGTACAAACACTGTCTCTGAAGTAAC	600		
Db 186826	AGTGTATAGTCTAGTGAAGTGACCTCAACTGTGTACAAACACTGTCTCTGAAGTAAC	186767		
QY 601	TTTCTAGACACAATAGAGACTCCAGACCTGAAACCTCTTCCGCCAAGATCTTAAGAGC	660		
Db 186766	TTTCTAGACACAATAGAGACTCCAGACCTGAAACCTCTTCCGCCAAGATCTTAAGAGC	186707		
QY 661	TCCATCTCCACCCATCTCATCATCAAGAGCGGGTGAAGCGCTGGCGCTAGCAAAACA	720		
Db 186706	TCCATCTCCACCCATCTCATCATCAAGAGCGGGTGAAGCGCTGGCGCTAGCAAAACA	186647		
QY 721	AATGAATCTGTGAGTGAGCGCCGAAAGGCTTTATGTATTCAGAAACACAAATCAAAAT	780		
Db 186646	AATGAATCTGTGAGTGAGCGCCGAAAGGCTTTATGTATTCAGAAACACAAATCAAAAT	186587		
QY 781	CCTCAGGAG 789			
Db 186586	CCTCAGGAG 186578			
RESULT 11				
LOCUS	AP001905			
DEFINITION	homo sapiens chromosome 18 clone RP11-723J4 map 18q12, WORKING DRAFT SEQUENCE, 23 unordered pieces.			
ACCESSION	AP001905			
VERSION	AP001905.3	GI:9188521		
KEYWORDS	HTG; HTGS_PHASE1; HTGS_DRAFT.			
SOURCE	Homo sapiens DNA, clone:RP11-723J4.			
ORGANISM	Homo sapiens			
	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.			

* 42545 64659: contig of 22115 bp in length
* 64660 64759: gap of 100 bp
* 64760 82476: contig of 18117 bp in length
* 82477 82976: gap of 100 bp
* 82977 100906: contig of 17930 bp in length
* 100907 101006: gap of 100 bp
* 101007 117951: contig of 16945 bp in length
* 117952 118051: gap of 100 bp
* 118052 131503: contig of 13452 bp in length
* 131504 131603: gap of 100 bp
* 131604 141537: contig of 9934 bp in length
* 141538 141637: gap of 100 bp
* 141638 152889: contig of 11252 bp in length
* 152890 152989: gap of 100 bp
* 152990 161818: contig of 8829 bp in length
* 161819 161918: gap of 100 bp
* 161919 170983: contig of 9065 bp in length
* 170984 171083: gap of 100 bp
* 171084 178239: contig of 7156 bp in length
* 178240 178339: gap of 100 bp
* 178340 187572: contig of 9233 bp in length
* 187573 187672: gap of 100 bp
* 187673 193032: contig of 5360 bp in length
* 193033 193132: gap of 100 bp
* 193133 198842: contig of 5710 bp in length
* 198843 198942: gap of 100 bp
* 198943 203032: contig of 4090 bp in length
* 203033 203132: gap of 100 bp
* 203133 207783: contig of 4651 bp in length
* 207784 207883: gap of 100 bp
* 207884 211428: contig of 3545 bp in length
* 211429 211528: gap of 100 bp
* 211529 215591: contig of 4063 bp in length
* 215592 215691: gap of 100 bp
* 215692 217832: contig of 2141 bp in length
* 217833 217932: gap of 100 bp
* 217933 220510: contig of 2578 bp in length
* 220511 220610: gap of 100 bp
* 220611 222035: contig of 1425 bp in length
* 222036 222135: gap of 100 bp
* 222136 223655: contig of 1520 bp in length
* 223656 223755: gap of 100 bp
* 223756 224788: contig of 1033 bp in length.
FEATURES
source
1. .224788
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="18"
/map="18q12"
/clone="RP11-723J4"
1. .42444
/note="assembly_fragment"
42545. .64659
/note="assembly_fragment clone_end:SP6 vector_side:right"
64760. .82976
/note="assembly_fragment"
82977. .100906
/note="assembly_fragment"
101007. .117951
/note="assembly_fragment"
118052. .131503
/note="assembly_fragment"
131604. .141537
/note="assembly_fragment clone_end:77 vector_side:right"
141638. .152889
/note="assembly_fragment"
152990. .161818
/note="assembly_fragment"
161919. .170983
/note="assembly_fragment"
171084. .178239
/note="assembly_fragment"
178340. .187572

misc_feature /note="assembly_fragment"
187673. .193032
misc_feature /note="assembly_fragment"
193133. .198842
misc_feature /note="assembly_fragment"
198943. .203032
misc_feature /note="assembly_fragment"
203133. .207783
misc_feature /note="assembly_fragment"
207884. .211428
misc_feature /note="assembly_fragment"
211529. .215591
misc_feature /note="assembly_fragment"
215692. .217832
misc_feature /note="assembly_fragment"
217933. .220510
misc_feature /note="assembly_fragment"
220611. .222035
misc_feature /note="assembly_fragment"
222136. .223655
misc_feature /note="assembly_fragment"
223756. .224788
/note="assembly_fragment"
BASE COUNT 67462 a 44418 c 45111 g 65596 t 2201 others
ORIGIN
Query Match 34.7% Score 787.4; DB 2; Length 224788;
Best Local Similarity 99.9% Pred. No. 3.7e-184;
Matches 788; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
Qy 1 ATGCGAGGAAGTTATCTGTAATCTTGATCTGACCTTGGCCCTCTCTGTCACAAATCCC 60
Db 2349 ATGCGAGGAAGTTATCTGTAATCTTGATCTGACCTTGGCCCTCTCTGTCACAAATCCC 2408
Qy 61 CTTTCATGAACCTAAAGCAGCTGCTTTCCCCAGACCACTGAGAAATTAAGTCCGAATTGG 120
Db 2409 CTTTCATGAACCTAAAGCAGCTGCTTTCCCCAGACCACTGAGAAATTAAGTCCGAATTGG 2468
Qy 121 GAATCTGGCATTATGTTGACTTGGCAATTTCCACAGCGCAATATCATCTACACAGCTT 180
Db 2469 GAATCTGGCATTATGTTGACTTGGCAATTTCCACAGCGCAATATCATCTACACAGCTT 2528
Qy 181 TTCTACCGCTATGAGAAATAATTTCTTGTGCTTCAAGGGTTTCAGAAAATTAATTCTCAA 240
Db 2529 TTCTACCGCTATGAGAAATAATTTCTTGTGCTTCAAGGGTTTCAGAAAATTAATTCTCAA 2588
Qy 241 AATATAGGCATAGATAAGATTTAAAGAAATCCATATACACCATGACCAACCATCACTCA 300
Db 2589 AATATAGGCATAGATAAGATTTAAAGAAATCCATATACACCATGACCAACCATCACTCA 2648
Qy 301 GACCACGAGCATCACTCAGACCATGAGCGTCACTCAGACCATGAGCATCACTCAGACCAC 360
Db 2649 GACCACGAGCATCACTCAGACCATGAGCGTCACTCAGACCATGAGCATCACTCAGACCAC 2708
Qy 361 GAGCATCACTCTGACCATGATCATCACTCTCAGCAATAATCATGCTGCTTCTGGTAAAAAT 420
Db 2709 GAGCATCACTCTGACCATGATCATCACTCTCAGCAATAATCATGCTGCTTCTGGTAAAAAT 2768
Qy 421 AAGGAAAAAGCTCTTTGCCCCAGACCATGACTCAGATAGTTTCAGGTAAAGATCCTAGAAAC 480
Db 2769 AAGGAAAAAGCTCTTTGCCCCAGACCATGACTCAGATAGTTTCAGGTAAAGATCCTAGAAAC 2828
Qy 481 AGCCAGGGGAAAGAGCTCAACGACCAAGAAATGCGCCAGTGGTAGAAGGAATGTCAAGGAC 540
Db 2829 AGCCAGGGGAAAGAGCTCAACGACCAAGAAATGCGCCAGTGGTAGAAGGAATGTCAAGGAC 2888
Qy 541 AGTGTAGTCTAGTGAAGTGACCTCAACTGTGTACAACACTGTCTCTCTGAAGGAACCTCAC 600
Db 2889 AGTGTAGTCTAGTGAAGTGACCTCAACTGTGTACAACACTGTCTCTCTGAAGGAACCTCAC 2948
Qy 601 TTTTCTAGACAAATAGAGACTTCCAGACCTGGAAAAACTCTTCCCCAAAGATGTGAAGCAGC 660
Db 2949 TTTTCTAGACAAATAGAGACTTCCAGACCTGGAAAAACTCTTCCCCAAAGATGTGAAGCAGC 3008

```
QY 661 TCCACTCCAGCAGCTGTACATCAAGAGCGGGTGAGCGGCTGTGCTGTAGCAAAACA 720
Db 3009 TCCACTCCAGCAGCTGTACATCAAGAGCGGGTGAGCGGCTGTGCTGTAGCAAAACA 1068
QY 721 AATGAATCTGTGAGTGAAGCCCGGAAAGGCTTTATGTATTCCAGAAACACAAATGAAAT 780
Db 3069 AATGAATCTGTGAGTGAAGCCCGGAAAGGCTTTATGTATTCCAGAAACACAAATGAAAT 1128
QY 781 CCTCAGCAG 789
Db 3129 CCTCAGCAG 3137

RESULT 12
AX465590
LOCUS AX465590 899 bp DNA linear PAT 16-JUL-2002
DEFINITION Sequence 3 from Patent WO0216939.
ACCESSION AX465590
VERSION AX465590.1 GI:21899893
KEYWORDS Mus sp.
SOURCE Mus sp.
ORGANISM Mus sp.
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus.
REFERENCE 1
AUTHORS Mack,D., Gish,K.C. and Wilson,K.E.
TITLE Methods of diagnosis of cancer and screening for cancer modulators
JOURNAL Patent: WO 0216939-A 3 28-FEB-2002;
FOS Biotechnology, Inc. (US)
FEATURES
    location/Qualifiers
        1..899
        /organism="Mus sp."
        /db_xref="taxon:10095"
        /note="prostate/breast/ovarian/bladder cancer protein
        mouse BCR4 CDNA"
        4..5897
        /note="mouse BCR4"
        /codon_start=1
        /protein_id="CAD42375.1"
        /db_xref="GI:21899894"
        /translation="MSIQALLNATFNLYLCPALINLIARACLIHTASEKKAEIPPKT
        YSLQIANLGGFIAISLISFLSLIGVILVPLMNRVFFKFLSFLVALAVGTISGDALLH
        LIPHSIAHQSHSHEEPAMEKKRGLPETHLSAONTESSEYFDSTWKGTLAIGLYFM
        FLVEHIVLTIKQDKKXXXXKKPENDESVSKQLSKYSQLSNFKVDPGRPPES
        YLRADSPSPSPDQOPDVGRRXDRHSDPCTPHKKSTNNMVRPGIQEQVPPFTLSRYAG
        PVRRPXHHHDYHHILXHLTR"
        PVRPXHHHDYHHILXHLTR"

BASE COUNT 252 a 239 c 198 g 203 t 7 others
ORIGIN
Query Match 27.3%; Score 619.6; DB 6; Length 899;
Best Local Similarity 84.3%; Pred. No. 8.9e-143;
Matches 752; Conservative 1; Mismatches 131; Indels 8; Gaps 5;

QY 823 GGCATGGCATCCAGTCCCGCTGAATGCAACAGAGTTCAACTATCTCTGTCACGCCATC 882
Db 1 GGCATGGCATCCAGTCTGTGTAATGCAACGAAATTAATCTCTCTGTCACGCCATC 60
QY 883 ATCAACCAATTCATGCTAGATCTTGTCTGATTTCATA---CAAGTGAAGAAAGCTGAA 939
Db 61 ATCAATCAATTCATGCTCGGGCTGTCTGATTTCATACAGCAAGTGAACAAAGCCAGAA 120
QY 940 ATCCCTCCAAAGACCTATTTCATACAAATAGCCTGGGTGTGGTTTATAGCCATTTC 999
Db 121 ATCCCTCCAAAGACCTATTTCATACAAATAGCCTGGGTGTGGTTTATAGCCATTTC 180
QY 1000 ATCATCAGTTTCCCTGCTCTGCTGGGGTTATCTTAGTGCCTCTCATCAATCCGCTGTTT 1059
Db 181 ATCATCAGTTTCCCTGCTCTGCTGGGATCATCTTGGTGCCCACTCATCAACCGCGTATT 240
QY 1060 TTCAAATTTCTCCTGAGTTTCTCTTGTGGCACTGGCGCTTGGGACATTTTCAGTGTGATGCT 1119
Db 241 TTCAAGTTCTCCTGAGCTTCTCTGCTGGGGCTGGCGCTCGGAACCTCAGTGGTATGCT 300
```

```
QY 1120 TTTTTCACACCTTTTCCACATTTCTCATGCAAGTCAACCACCATAGTATAGCCATGAAGA 1179
Db 301 CTGTTTACATCTTCTCCACACTCTCATGCAAGTCAATCAGCAGCTCATAGCCATGAAGAG 360
QY 1180 CCAGCAATGGAATCAAAAGAGGAGCACCTTTTTCAGTCTCATCTCTTCTCAAAAACATAGAA 1239
Db 361 CCAGCGATGGAATGAAAAGAGGCGCCCTGTTCACCCACCTGTGCGGCTCAGAAATATAGAA 420
QY 1240 GAAAGTGCTATTTTTCACGTGGAGGGTCTTAACAGGCTCTAGGAGGCTGTGATTTTC 1299
Db 421 GAAAGCTCTATTTTTCACGTGGAGGGTCTTAACAGGCTCTAGGAGGCTGTGATTTTC 480
QY 1300 ATGTTTCTTGTGTAACATGTCTCATATTGATCAAAACAATTTTAAAGATAAAGAAAAAG 1359
Db 481 ATGTTTCTTGTGTAACATGTCTCATATTGATCAAGCAATTTTAAAGATAAAGAAAAANA 540
QY 1360 AATCAAGAAAGACCTGAAATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1419
Db 541 AANMAAAGAAACCTGAAATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 600
QY 1420 GAATCTCAACTTTTCAACAAATGAGGAGAAAGTAGATACAGATGATGATGATGATGATGAT 1479
Db 601 GACTCTCAGCTTCTCTCAAAATGAAGAGAGCTGGACCCAGGGAACGACCTGAAAGCTAT 660
QY 1480 TTACGAGCAGACTCAAGAGCCCTCCACACTTTGATTCTCAGCAGCCCTG-CAGTCTTGG 1538
Db 661 CTGCGAGCGGACTCCCAAGAGCCCTCCACACTTTGATTCTCCACAGCCGAGGAGTGTGGA 720
QY 1539 AGAAGAAGAGTCTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 1595
Db 721 AGAAGAAGAGTCTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 780
QY 1596 ACCCAGAGGGTGCAGAAATAAATGCCATTACATTTCCACGATACACTCGGCCAGTCCAGA 1655
Db 781 CCCAGGGGGTGCAGAAATAAATGCCATTACACTTTACAGTACGCTGGGCCAGTCCGA 840
QY 1656 CGATCTCATTCACGACCATCATGACTACCATCATCATTTCTCCATCATCACCAC 1707
Db 841 CGACCT-ATNCACCACCATCATGACTACCATCATTTCTGNACCACCACTC 891

RESULT 13
LOCUS AX207216 551 bp DNA linear PAT 30-AUG-2001
DEFINITION Sequence 12 from Patent WO0155178.
ACCESSION AX207216
VERSION AX207216.1 GI:15394970
KEYWORDS synthetic construct.
SOURCE synthetic construct.
ORGANISM artificial sequences.
REFERENCE 1 (bases 1 to 551)
AUTHORS Goddard,A., Gurney,A.L., Smith,V., Hongo,J.A. and de Sauvage,F.
TITLE Compositions and methods for treatment of cancer
JOURNAL Patent: WO 0155178-A 12 02-AUG-2001;
GENENTECH, INC. (US)
FEATURES
    source
        Location/Qualifiers
            1..551
            /organism="synthetic construct"
            /db_xref="taxon:32630"
            /note="CDNA"
BASE COUNT 128 a 130 c 127 g 166 t
ORIGIN
Query Match 24.2%; Score 549.4; DB 6; Length 551;
Best Local Similarity 99.8%; Pred. No. 2.3e-125;
Matches 550; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1618 TGCCATTTCACATTTCCACGATACACTCGGCCAGTCAGACCATCTCATTTACCACCATCAT 1677
Db 1 TGCCATTTCACATTTCCACGATACACTCGGCCAGTCAGACCATCTCATTTACCACCATCAT 60
```

```

QY 1678 GACTACCATCATATTCATCATCACCACCACCAAAACCCACCACTCTCACAGTCACAGC 1737
|||||
Db 61 GACTACCATCATATTCATCATCACCACCACCAAAACCCACCACTCTCACAGTCACAGC 120
|||||
QY 1738 CAGCGCTACTCTCGGAGGAGCTGAAGATGCCGCGCTCGGCACCTTTGGCCTGGATGCTG 1797
|||||
Db 121 CAGCGCTACTCTCGGAGGAGCTGAAGATGCCGCGCTCGGCACCTTTGGCCTGGATGCTG 180
|||||
QY 1798 ATAATGGGTGATGGCTGCACAAATTTACAGCATGGCCTAGCAATTTGGTGGCTTTTACT 1857
|||||
Db 181 ATAATGGGTGATGGCTGCACAAATTTACAGCATGGCCTAGCAATTTGGTGGCTTTTACT 240
|||||
QY 1858 GAAGGCTTATCAAGTGGTAAAGTACTTCTGCTGTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1917
|||||
Db 241 GAAGGCTTATCAAGTGGTAAAGTACTTCTGCTGTGCTGCTGCTGCTGCTGCTGCTGCTGCT 300
|||||
QY 1918 GAATTTAGGTGACTTTGCTGTTCTACTAAAGGCTGGCATGACCGTTAAAGCAGGCTGCTGCT 1977
|||||
Db 301 GAATTTAGGTGACTTTGCTGTTCTACTAAAGGCTGGCATGACCGTTAAAGCAGGCTGCTGCT 360
|||||
QY 1978 TATAATGCATTTCTCAGCCATGCTGGCGTATCTTGGAAATGGCAACAGAAATTTTCATTTGGT 2037
|||||
Db 361 TATAATGCATTTCTCAGCCATGCTGGCGTATCTTGGAAATGGCAACAGAAATTTTCATTTGGT 420
|||||
QY 2038 CATTTATGCTGAAATGTTTCTATGTGATATTTGCACTTACTGCTGCTGCTGCTGCTGCTGCTGCT 2097
|||||
Db 421 CATTTATGCTGAAATGTTTCTATGTGATATTTGCACTTACTGCTGCTGCTGCTGCTGCTGCTGCT 480
|||||
QY 2098 GTTGTCTGCTGATATGATGCTACCTGAAATGCTGCACAAATGATGCTAGTACCATGGATGT 2157
|||||
Db 481 GTTGTCTGCTGATATGATGCTACCTGAAATGCTGCACAAATGATGCTAGTACCATGGATGT 540
|||||
QY 2158 AGCCGCTGGG 2168
|||||
Db 541 AGCCGCTGGG 551
|||||

RESULT 14
LOCUS AX017261
DEFINITION Sequence 12 from Patent WO947669.
ACCESSION AX017261
VERSION AX017261.1 GI:10042179
KEYWORDS
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominoidea; Homo.
REFERENCE 1 (bases 1 to 1597)
AUTHORS Schmitt,A., Specht,T., Dahl,E., Hinzmann,B., Rosenthal,A. and
Pilarczyk,C.
TITLE Human nucleic acid sequences from tissue of breast tumors
JOURNAL.
PATENT: WO 947669-A 12 23-SEP-1999;
SCHMITT ARMIN (DE); SPECHT THOMAS (DE); DNIIL ENCAR (DE); HINZMANN
BERND (DE); ROSENTHAL ANDRE (DE); METAGEN GES FUER GENOMFORSCHUN
(DE); PILARSKY CHRISTIAN (DE)
FEATURES
source
Location/Qualifiers
1..1597
/organism="Homo sapiens"
/db_xref="taxon:9606"
BASE COUNT 476 a 229 c 337 g 555 t
ORIGIN
Query Match 22.5% Score 509.4; DB 6; Length 1597;
Best local Similarity 99.8%; Pred. No. 2.2e-115;
Matches 510; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1758 GCTGAAGATGCGCGCTGCGCACTTTGGCTGGATGGTGAATAGGATGATGCTGCA 1817
|||||
Db 19 GCTGAAGATGCGCGCTGCGCACTTTGGCTGGATGGTGAATAGGATGATGCTGCA 78
|||||
QY 1818 CAATTTACGCGATGCGCTAGCAATTTGCTGCTTTTACTGAAGGCTTATCAAGTCTTTT 1877
|||||

```

```

Db 79 CAATTTACGCGATGGCGCTAGCAATTTGCTGCTGCTTTTACTGAAGGCTTATCAAGTGGTTT 138
QY 1878 AAGTACTTTCTGTTGCTGCTGCTGCTCATGAGTTGCCCTCATGAATTAGTGACTTTGCTGT 1937
|||||
Db 139 AAGTACTTTCTGTTGCTGCTGCTGCTCATGAGTTGCCCTCATGAATTAGTGACTTTGCTGT 198
|||||
QY 1938 TCTACTAAAGGCTGGCATGACCGTTAAAGCAGGCTGCTCTTTTATAATGCAATTTGCAAGCAT 1997
|||||
Db 199 TCTACTAAAGGCTGGCATGACCGTTAAAGCAGGCTGCTCTTTTATAATGCAATTTGCAAGCAT 258
|||||
QY 1998 GCTGGCGCTATCTTGGAAATGCAACAGCAATTTTCATTTGGTCAATTTGCTGCAAAATGTTTC 2057
|||||
Db 259 GCTGGCGCTATCTTGGAAATGCAACAGCAATTTTCATTTGGTCAATTTGCTGCAAAATGTTTC 318
|||||
QY 2058 TATGTGGAATTTGCACTTACTGCTGCTGCTTATTCATGATGATGCTGCTGCTGCTGCTGCTGCT 2117
|||||
Db 319 TATGTGGAATTTGCACTTACTGCTGCTGCTTATTCATGATGATGCTGCTGCTGCTGCTGCTGCT 378
|||||
QY 2118 ACCTGAATGCTGCACAAATGATGCTAGTACCATGATGATGATGATGATGATGATGATGATGAT 2177
|||||
Db 379 ACCTGAATGCTGCACAAATGATGCTAGTACCATGATGATGATGATGATGATGATGATGATGAT 438
|||||
QY 2178 TTTACAGAAATGCTGGGATGCTTTTGGGTTTGGCAATTTACTTATTTTCCATATTTTCA 2237
|||||
Db 439 TTTACAGAAATGCTGGGATGCTTTTGGGTTTGGCAATTTACTTATTTTCCATATTTTCA 498
|||||
QY 2238 ACATAAAATGCTGCTTTCGTATAAAATTTCTAG 2268
|||||
Db 499 ACATAAAATGCTGCTTTCGTATAAAATTTCTAG 529
|||||

RESULT 15
LOCUS AC126667
DEFINITION 193437 bp DNA linear HTG 11-JUL-2002
Rattus norvegicus clone CH230-5D17, *** SEQUENCING IN PROGRESS ***
69 unordered pieces.
ACCESSION AC126667
VERSION AC126667.2 GI:21717863
KEYWORDS HTG; HTGS_PHASE1.
SOURCE Norway rat.
ORGANISM Rattus norvegicus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
Rattus.
REFERENCE 1 (bases 1 to 193437)
AUTHORS Muzny,D.M., Adams,C., Adio-Oduola,B., Ali-osman,F.R., Allen,C.,
Alsbrooks,S.L., Amaratunga,H.C., Are,J.R., Ayale,M., Banks,T.,
Barbaria,J., Benton,J., Blimage,K., Blankenburg,K., Bounin,D.,
Buck,J., Bowle,S., Brieva,M., Brown,E., Brown,M., Bryant,N.P.,
Buhay,C., Burch,P., Burkett,C., Burrell,K.L., Byrd,N.C.,
Carroll,T.F., Carter,M., Cavazos,S.K., Chacko,J., Chavez,D.,
Chen,G., Chen,R., Chen,Z., Chowdhry,I., Christopoulos,C.,
Cleveland,C.D., Cox,C., Coyle,M.D., Dathorne,S.R., David,R.,
Delaney,K.R., Delgado,O., Denn,A.L., Ding,Y., Dinh,H.H.,
Douthwaite,K.J., Draper,H., Dugan-Rocha,S., Durbin,K.J.,
Earnhart,C., Edgar,D., Edwards,C.C., Elhaj,C., Escotto,M.,
Falls,T., Ferraguto,D., Flagg,N., Ford,J., Foster,P., Frantz,P.,
Gabisi,A., Gao,J., Garcia,A., Garner,T., Garza,N., Gill,R.,
Gorrell,J.H., Guevara,W., Gunaratne,P., Hale,S., Hamilton,K.,
Harris,C., Harris,K., Hart,M., Havlak,P., Hawes,A., Hernandez,J.,
Hernandez,O., Hodgson,A., Hogue,M., Holloway,C., Hollins,B.,
Honsi,F., Howard,S., Huber,J., Hulyk,S., Hume,J., Jackson,L.E.,
Jacobson,B., Jia,Y., Johnson,R., Jollivet,S., Joudah,S.,
Karlssoon,E., Kelly,S., Khan,U., King,L., Korval,J., Kovar,C.,
Kratovic,J., Kureshi,A., Landry,N., Leal,B., Lewis,L.C., Lewis,L.,
Li,J., Li,Z., Lichtarge,O., Lieu,C., Liu,J., Liu,W., Louised,H.,
Lozaro,R.J., Lu,X., Lucier,A., Lucier,R., Luna,R., Ma,J.,
Maheshwari,M., Mapua,P., Martin,R., Martindale,A., Martinez,E.,
Massey,G., Mawhinney,E., McLeod,M.P., Meador,M., Mei,G., Metzger,M.,
Miner,G., Miner,Z., Mitchell,T., Mohabbat,K., Morgan,M., Morris,S.,
Moser,M., Neal,D., Newton,J., Newton,N., Nguyen,A., Nguyen,N.,
Nguyen,N., Nickerson,E., Nwokenkwo,S., Oguh,M., Okwionu,G.,

```


Oraqunye, N., Oviedo, R., Pace, A., Payton, B., Peery, J., Perez, L.,
 Peters, L., Pickens, K., Primus, E., Pu, L.L., Quiles, M., Ken, Y.,
 Rives, M., Rojas, A., Rojubokan, I., Rolfe, M., Ruiz, S., Savary, G.,
 Scherer, S., Scott, G., Shen, H., Shooshitari, N., Sisson, I.,
 Sodergren, E., Sonaik, T., Sparks, A., Stanley, H., Stone, H.,
 Sulton, A., Svatek, A., Tabor, P., Tamerisa, A., Tamerisa, K., Tang, H.,
 Tansey, J., Taylor, C., Taylor, T., Telford, B., Thomas, N., Thomas, S.,
 Usmani, K., Vasquez, L., Vera, V., Villalón, D., Vinson, R., Wang, Q.,
 Wang, S., Ward-Moore, S., Warren, R., Washington, C., Watlington, S.,
 Williams, G., Williamson, A., Wleczek, R., Wooden, S., Worley, K.,
 Wu, C., Wu, Y., Wu, Y.F., Zhou, J., Zorrilla, S., Nelson, D.,
 Weinstein, G., and Gibbs, R.
 Direct Submission
 Unpublished
 2 (bases 1 to 193437)
 Worley, K.C.
 Direct Submission
 Submitted (08-JUL-2002) Human Genome Sequencing Center, Department
 of Molecular and Human Genetics, Baylor College of Medicine, One
 Baylor Plaza, Houston, TX 77030, USA
 3 (bases 1 to 193437)
 Worley, K.C.
 Direct Submission
 Submitted (11-JUL-2002) Human Genome Sequencing Center, Department
 of Molecular and Human Genetics, Baylor College of Medicine, One
 Baylor Plaza, Houston, TX 77030, USA
 On Jul 10, 2002 this sequence version replaced qi:21703549.
 ----- Genome Center
 Center: Baylor College of Medicine
 Center code: BCM
 Web site: <http://www.hgsc.bcm.tmc.edu/>
 Contact: hgsc-help@bcm.tmc.edu
 ----- Project Information
 Center project name: GBB1
 Center clone name: CH230-5p17
 ----- Summary Statistics
 Sequencing vector: Plasmid;
 Chemistry: Dye-terminator Big Dye; 100% of reads
 Assembly program: Phrap; version 0.990329
 Consensus quality: 143517 bases at least Q40
 Consensus quality: 150307 bases at least Q30
 Consensus quality: 154257 bases at least Q20

 * NOTE: Estimated insert size may differ from sequence length
 * (see http://www.hgsc.bcm.tmc.edu/docs/genbank_draft_data.html).
 * NOTE: This is a 'working draft' sequence. It currently
 * consists of 69 contigs. The true order of the pieces
 * is not known and their order in this sequence record is
 * arbitrary. Gaps between the contigs are represented as
 * runs of N, but the exact sizes of the gaps are unknown.
 * This record will be updated with the finished sequence
 * as soon as it is available and the accession number will
 * be preserved.
 *
 * 1172: contig of 1172 bp in length
 * 1173 1272: gap of unknown length
 * 1273 2487: contig of 1215 bp in length
 * 2488 2587: gap of unknown length
 * 2588 4394: contig of 1807 bp in length
 * 4395 4494: gap of unknown length
 * 4495 5815: contig of 1321 bp in length
 * 5816 5915: gap of unknown length
 * 5916 7626: contig of 1711 bp in length
 * 7627 7726: gap of unknown length
 * 7727 9206: contig of 1480 bp in length
 * 9207 9306: gap of unknown length
 * 9307 10449: contig of 1143 bp in length
 * 10450 10549: gap of unknown length
 * 10550 11609: contig of 1060 bp in length
 * 11610 11709: gap of unknown length
 * 11710 12957: contig of 1248 bp in length
 * 12958 13057: gap of unknown length
 * 13058 14501: contig of 1444 bp in length
 * 14502 14601: gap of unknown length
 *
 * 14602 15762: contig of 1161 bp in length
 * 15763 15862: gap of unknown length
 * 15863 16877: contig of 1015 bp in length
 * 16878 16977: gap of unknown length
 * 16978 18761: contig of 1784 bp in length
 * 18762 18861: gap of unknown length
 * 18862 20537: contig of 1676 bp in length
 * 20538 22637: gap of unknown length
 * 22638 22730: contig of 2093 bp in length
 * 22731 22830: gap of unknown length
 * 22831 25262: contig of 2432 bp in length
 * 25263 25362: gap of unknown length
 * 25363 27262: contig of 1900 bp in length
 * 27263 27362: gap of unknown length
 * 27363 28524: contig of 1162 bp in length
 * 28525 28624: gap of unknown length
 * 28625 30233: contig of 1609 bp in length
 * 30234 30333: gap of unknown length
 * 30334 31883: contig of 1550 bp in length
 * 31884 33437: gap of unknown length
 * 33438 33536: contig of 1453 bp in length
 * 33537 34740: contig of 1203 bp in length
 * 34741 36917: contig of 2078 bp in length
 * 36918 38173: contig of 1156 bp in length
 * 38174 38273: gap of unknown length
 * 38274 39952: contig of 1679 bp in length
 * 39953 40052: gap of unknown length
 * 40053 42000: contig of 1948 bp in length
 * 42001 42100: gap of unknown length
 * 42101 43702: contig of 1602 bp in length
 * 43703 43802: gap of unknown length
 * 43803 46149: contig of 2347 bp in length
 * 46150 46249: gap of unknown length
 * 46250 47598: contig of 1349 bp in length
 * 47599 49210: contig of 1512 bp in length
 * 49211 49310: gap of unknown length
 * 49311 50780: contig of 1470 bp in length
 * 50781 50880: gap of unknown length
 * 50881 52850: contig of 1970 bp in length
 * 52851 52950: gap of unknown length
 * 52951 53981: contig of 1031 bp in length
 * 53982 54081: gap of unknown length
 * 54082 55921: contig of 1740 bp in length
 * 55922 57909: contig of 1988 bp in length
 * 57910 58009: gap of unknown length
 * 58010 60388: contig of 2379 bp in length
 * 60389 60488: gap of unknown length
 * 60489 62460: contig of 1972 bp in length
 * 62461 62560: gap of unknown length
 * 62561 64500: contig of 1940 bp in length
 * 64501 64600: gap of unknown length
 * 64601 66889: contig of 2289 bp in length
 * 66890 66989: gap of unknown length
 * 66990 69589: contig of 2600 bp in length
 * 69590 71653: contig of 1964 bp in length
 * 71654 71753: gap of unknown length
 * 71754 75273: contig of 3520 bp in length
 * 75274 75373: gap of unknown length
 * 75374 77772: contig of 2399 bp in length
 * 77773 77872: gap of unknown length
 * 77873 81413: contig of 3541 bp in length
 * 81414 81513: gap of unknown length
 * 81514 85249: contig of 3736 bp in length
 * 85250 85349: gap of unknown length
 * 85350 87366: contig of 2017 bp in length
 * 87367 87466: gap of unknown length
 * 87467 90125: contig of 2659 bp in length

```
* 90126 90225: gap of unknown length
* 90226 93089: contig of 2864 bp in length
* 93189 93190: gap of unknown length
* 93190 96456: contig of 3267 bp in length
* 96457 96556: gap of unknown length
* 96557 99751: contig of 3195 bp in length
* 99752 99851: gap of unknown length
* 99852 103352: contig of 3501 bp in length
* 103353 103452: gap of unknown length
* 103453 107023: contig of 3571 bp in length
* 107024 107123: gap of unknown length
* 107124 112368: contig of 5245 bp in length
* 112369 112469: gap of unknown length
* 112469 115712: contig of 3244 bp in length

Query Match 18.6%; Score 421.4; DB 2; Length 19,437;
Rest Local Similarity 75.7%; Pred. No. 2.6e-93;
Matches 606; Conservative 0; Mismatches 126; Indels 69; Gaps 4;

QY 1 ATGGCGAGGAAGTTATCTGTAATCTTGATCTGTGACCTTTGGCCCTCTCTGTGCACAAATGCC 60
    |||||  ||  |||||  |||||  |||||  |||||  |||||  ||  |||||  |||||
Db 107530 ATGGCCACAATTTATCTGTAATCATGATCTTGACCTTTGGCCCTTACAAATCCC 107589

QY 61 CTTTCATGAACATAA---AGCAGCTGCTTTCCCCAGACCACTGAGAAAATTAGTCCGAATP 117
    |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
Db 107590 CTTTCATGAACATAAATCAACAGCTGCTTTCTCTCAGACCACTGAGAAAATTAAATCAAAAT 107649

QY 118 TCGGAATCTGGCATTTAATCTTGACCTTGCAATTTCCACACGGCAATATCATCTACAACAG 177
    |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
Db 107650 TCGGAATCTGGCATTTAATCTTGACCTTGCAATTTCCACACGGCAATATCATCTACAACAG 107709

QY 178 CTTTTCATGCGGTATGGAGAAAATAATTTTGTGAGTTGAAGGTTTCAGAAAATTACTT 237
    |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
Db 107710 CTTTTCATGCGGTATGGAGAAAATAATTTTGTGAGTTGAAGGTTTCAGAAAATTACTT 107769

QY 238 CAAAATATAGCATAGATAAGATTAAGAATCCATATACACCATGACCCAGCCATACAC 297
    |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
Db 107770 CAGAATATAGCATAGATAAGATTAAGAATCCATATACACCATGACCCAGCCATACAC 107820

QY 298 TCAGACCCAGGACATCACTCAGACCATGAGCGTCACTCAGACCATGAGCATCTCAGAC 357
    |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
Db 107821 -----GAGCGTCACTCTGAC 107835

QY 358 CACGAGCATCACTCGACCATGATCATCACTCTCACCATAATCATGCTGCTTCTGGTAAA 417
    |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
Db 107836 CACGAGCCACCTCTGACCCAGGCGGCACTCCACCGAGGTCACTGCTGCCGCTGGGAAA 107895

QY 418 AATAAGCCAAAAGCTTTTGGCCAGACCATCACTCAGATAGTTTCAGGTAAGATCTTAGA 477
    |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
Db 107896 AACAGTCCGAAGGCTTTTGGCCAGACCTTGACTCTGATAATTCAGGTAAATAATCCGAAC 107955

QY 478 AACAGTCCAGGGGAAGGAGCTTCAGCCAGACAGATGCCAGTGGTAGAAGGAATGTCAAG 537
    |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
Db 107956 ACCAGTCAAGCTTAAGGATCTCGCCAGCAGGATGTGAATGCGAGGAGGAATGGCAAG 108015

QY 538 TACAGTGTGTAGTGTAGTGAAGTGAACCTGTGTACAACTGTCTCTGAAGGAAC 597
    |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
Db 108016 CAGAGTGAAGGCTCTAGTGAAGTGAACCTGTGTACAACTGTCTCTGAAGGAAC 108075

QY 598 CACTTTCTAGACAGAAATGACACTCCAGACCTGGAATACTCTTCCCAAGAGATCTAAGC 657
    |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
Db 108076 CACTTTCTAGACAGAAATGACACTCCAGACCTGGAATACTCTTCCCAAGAGATCTAAGC 108132

QY 658 AGCTCCACTCCACCACTGTGCATCAATAAGAGCGGGTGGCCCGGCTG-----GCT 708
    |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
Db 108133 CTTTCTACCCCACTCAGCATCACAGAGAGAGCGCACTGGCGCGCTGAGTCCAGCTAGCT 108192

QY 709 GTAGGAAAACAAATGATCTGTGAGTGAAGTCCCGCAAGCGCTTTATGTATTTCAAGAAC 768
    |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
Db 108193 AGGAGGAAGCAACGATTTGTGAGTGAAGTCCCGCAAGCGCTTTATGTATTTCAAGAAC 108252

QY 769 ACAAATGAAATCTCTCAGGAG 789
    |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
```

Db 108253 TCAAATGACAATATATCCAGGAG 108273

Search completed: November 28, 2002, 04:15:24
Job time : 6214 secs

GenCore version 5.1.3
Copyright (c) 1993 - 2002 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run On: November 27, 2002, 23:26:15 ; Search Time 2851 Seconds

(without alignments)
12883.677 Million cell updates/sec

Title: US-09-642-034-4

Perfect score: 2268

Sequence: 1 atggcaggaggtatctgt.....tgttcyatalaaatttctay 2268

Scoring table: IDENTITY_NUC

Gapop 10.0 , Gapext 1.0

Searched: 16154066 seqs, 8097743376 residues

Total number of hits satisfying chosen parameters: 32308132

Minimum DB seq length: 0

Maximum DB seq length: 20000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

EST:*

1: em_estba:*

2: em_esthm:*

3: em_estin:*

4: em_estnu:*

5: em_estov:*

6: em_estpl:*

7: em_estro:*

8: em_hlc:*

9: gb_est1:*

10: gb_est2:*

11: gb_hlc:*

12: gb_est3:*

13: gb_est4:*

14: gb_est5:*

15: em_estfun:*

16: em_estom:*

17: gb_gss:*

18: em_gss_hum:*

19: em_gss_inv:*

20: em_gss_pln:*

21: em_gss_vrt:*

22: em_gss_fun:*

23: em_gss_mam:*

24: em_gss_mus:*

25: em_gss_other:*

26: em_gss_pro:*

27: em_gss_rod:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	877.6	38.7	1074	13	BM480018
2	799.6	35.3	872	9	AU120027
3	724	31.9	810	13	BM456317
4	692	30.5	701	14	BQ027619
5	652	28.7	1135	13	BM560789
6	649.8	28.7	674	14	BM978669

7 627.6 27.7 12 BG168169
8 600.2 26.5 12 BG530601
9 597.6 26.3 14 BQ581865
10 586.8 25.9 766 14 BM948026
11 576 25.4 925 12 BF983458
12 571.4 25.2 743 9 AI907176
13 567.4 25.0 829 9 AU079709
14 566.8 25.0 751 14 BM963737
15 566.2 25.0 887 13 BI150324
16 564 24.9 781 14 BM949686
17 563.8 24.9 857 12 BF032013
18 538.6 23.7 948 14 BQ955836
19 526.6 23.2 580 12 BF915495
20 517.8 22.8 849 12 BF680209
21 515.8 22.7 791 12 BG547794
22 505.2 22.3 682 14 BM949679
23 503.4 22.2 950 14 BQ433766
24 502.2 22.1 518 10 AW178411
25 501.8 22.1 513 10 AW178409
26 499.6 22.0 561 9 AL042316
27 499.2 22.0 1003 13 BM478134
28 495.8 21.9 560 12 BF915709
29 493.6 21.8 500 10 AW178317
30 493.4 21.8 495 10 AW178381
31 493.4 21.8 522 10 AW383696
32 493.4 21.8 522 10 AW383720
33 489.6 21.6 518 10 AW383693
34 489.6 21.6 518 10 AW383717
35 481.6 21.2 1112 14 BQ932626
36 475 20.9 593 10 AW365994
37 474 20.9 933 13 BI861563
38 472.4 20.8 507 12 BF699194
39 472.2 20.7 739 9 AA666965
40 469 20.7 694 9 AA666964
41 459.8 20.3 605 13 BI284823
42 458 20.2 1059 13 BM557996
43 454 20.0 528 9 AA242758
44 453.2 20.0 501 10 AW994833
45 446 19.7 484 13 BG994120

ALIGNMENTS

RESULT 1
BM480018
LOCUS BM480018 1074 bp mRNA linear EST 05-FEB-2002
DEFINITION AGENCOURT_6424145 NIH_MGC_67 Homo sapiens cDNA clone IMAGE:5491572
5', mRNA sequence.
ACCESSION BM480018.1 GI:18529060
VERSION BM480018
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 1074)
AUTHORS NIH-MGC <http://mgc.nci.nih.gov/>.
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL Unpublished (1997)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgapbs-r@mail.nih.gov
Tissue Procurement: ATCC
CDNA Library Preparation: Life Technologies, Inc.
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Agencourt Bioscience Corporation
Clone Distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at:
<http://image.llnl.gov>
Plate: LLAM2112 row: e column: 13
High quality sequence stop: 652.
Location/Qualifiers
1. .1074


```
QY 1536 GGAAGAGAGAGGTCATGATAGCTCATGCTCATCCACAGGAAGTCTACAAATTAATGT 1595
DB 1537 GGAAGAGAGAGGTCATGATAGCTCATGCTCATCCACAGGAAGTCTACAAATTAATGT 300
QY 1596 ACCAGAGGGTGCAGGAATAAATGCCATTTACATTTCCAGATACATTCGCGCCAGTACAGA 1655
DB 1656 ACCAGAGGGTGCAGGAATAAATGCCATTTACATTTCCAGATACATTCGCGCCAGTACAGA 360
QY 1656 CGATCTCATTCACCAACCATGATGACTACCATCATATTTCCATCATCACCACCAACAAA 1715
DB 1716 CGATCTCATTCACCAACCATGATGACTACCATCATATTTCCATCATCACCACCAACAAA 420
QY 1716 CCACATCTCTCACAGTCCAGCGCTACTCTCGGAGGAGCTGAAGATGCGCGCT 1775
DB 1775 CCACATCTCTCACAGTCCAGCGCTACTCTCGGAGGAGCTGAAGATGCGCGCT 480
QY 1776 CGCCACATTTGGCGTGGATGATTAATGGGTGATGGCTGCACAAATTTACAGCATGCGCT 1835
DB 1835 CGCCACATTTGGCGTGGATGATTAATGGGTGATGGCTGCACAAATTTACAGCATGCGCT 540
QY 1836 AGCAATTTGCTGCTGCTTTACTGAAGCTTATCAAGTGGTTTAAGTACTTCTGTTGCTGT 1895
DB 1895 AGCAATTTGCTGCTGCTTTACTGAAGCTTATCAAGTGGTTTAAGTACTTCTGTTGCTGT 600
QY 1896 GTTCTGTCATGAGTTGCTCTCATTAATTAAGTGACTTTTCTCTACTAAAGGCTGCGCAT 1955
DB 1955 GTTCTGTCATGAGTTGCTCTCATTAATTAAGTGACTTTTCTCTACTAAAGGCTGCGCAT 660
QY 1956 GACCTTTAAGCAGGCTGTCCTTTTATTAATGATGCTGAGCCATGCTGCGGTATCTTGAAT 2015
DB 2015 GACCTTTAAGCAGGCTGTCCTTTTATTAATGATGCTGAGCCATGCTGCGGTATCTTGAAT 720
QY 2016 GCAACAGGAATTTTCTGCTATGCTGAAATGTTTCTATGATGGATATTTGCACT 2075
DB 2075 GCAACAGGAATTTTCTGCTATGCTGAAATGTTTCTATGATGGATATTTGCACT 780
QY 2076 TACTGCTGCTGCTATGATGATGTTG--CTCTGCTGATATGTTGCTGAAATGCTGCACT 2133
DB 2133 TACTGCTGCTGCTATGATGATGTTG--CTCTGCTGATATGTTGCTGAAATGCTGCACT 840
QY 2134 AAT--GATGCTAGTACCATGG 2153
DB 2153 AAT--GATGCTAGTACCATGG 861

RESULT 3
LOCUS BM456317 810 bp mRNA linear EST 05-FEB-2002
DEFINITION AGENCOURT_6409853 NIH_MGC_85 Homo sapiens cDNA clone IMAGE:5496678
5', mRNA sequence.
ACCESSION BM456317
VERSION BM456317.1 GI:18505357
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 810)
NIH-MGC http://mgi.nci.nih.gov/.
National Institutes of Health, Mammalian Gene Collection (MGC)
Unpublished (1999)
Contact: Robert Strausberg, Ph.D.
Email: cqaaps-r@mail.nih.gov
Tissue Procurement: Lou Staudt
cDNA Library Preparation: Life Technologies, Inc.
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Agencourt Bioscience Corporation
Clone Distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
Plate: LLAM12125 row: j column: 07
High quality sequence stop: 640.
```

```
FEATURES
Source
Location/Qualifiers
1..810
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:5496678"
/tissue_lib="NIH_MGC_85"
/tissue_type="lymphoma, cell line"
/lab_host="DH10B (phage-resistant)"
/notes="Organ: lymph; Vector: pCMV-SPORT6; Site_1: NotI;
Site_2: SalI; Cloned unidirectionally; oligo-dT primed.
Average insert size 1.867 kb. Library enriched for
full-length clones and constructed by Life Technologies.
Note: this is a NIH_MGC Library."
1 others
BASE COUNT 246 a 172 c 165 g 226 t
ORIGIN
Query Match 31.9%; Score 724; DB 13; Length 810;
Best Local Similarity 96.2%; Pred. No. 1.1e-151;
Matches 764; Conservative 0; Mismatches 25; Indels 5; Gaps 2;
QY 866 ATCTCTGTCAGCCATCATCAACCAATTTGATGCTAGATCTTGTCTGATTCATACAAATG 925
DB 1 ATCTCTGTCAGCCATCATCAACCAATTTGATGCTAGATCTTGTCTGATTCATACAAATG 60
QY 926 AAAAGAGGCTGAAATCCCTCCAAAGACCTATTTCATTACAAATAGCCTGGGTTGGTGTT 985
DB 61 AAAAGAGGCTGAAATCCCTCCAAAGACCTATTTCATTACAAATAGCCTGGGTTGGTGTT 120
QY 986 TTATAGCCATTTCCATCATCAGTTTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCA 1045
DB 121 TTATAGCCATTTCCATCATCAGTTTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCA 180
QY 1046 TGAATCGGCTGTTTTCAAATTTCTCCCTGAGTTTCTCTGTCGACCTGGCCGTTGGACTT 1105
DB 181 TGAATCGGCTGTTTTCAAATTTCTCCCTGAGTTTCTCTGTCGACCTGGCCGTTGGACTT 240
QY 1106 TGAGTGGTGTATGCTTTTACACCTTTCTTCCACATTTCTCCTCAAGTTCACCACTAGTC 1165
DB 241 TGAGTGGTGTATGCTTTTACACCTTTCTTCCACATTTCTCCTCAAGTTCACCACTAGTC 300
QY 1166 ATAGCCATGAAGACAGCAATGGAAATGAAAGAGGACCACTTTTTCAGTCTATCTGTCTT 1225
DB 301 ATAGCCATGAAGACAGCAATGGAAATGAAAGAGGACCACTTTTTCAGTCTATCTGTCTT 360
QY 1226 CTCAAACATAGAAAGAGTGCCTATTTTGATTCACGCTGGGAAGGCTTAACAGCTCTAG 1285
DB 361 CTCAAACATAGAAAGAGTGCCTATTTTGATTCACGCTGGGAAGGCTTAACAGCTCTAG 420
QY 1286 GAGGCTGTATTTTCATGTTTCTTGTGTAACATGCTCTCATTGATCAAAACAATTTAAAG 1345
DB 421 GAGGCTGTATTTTCATGTTTCTTGTGTAACATGCTCTCATTGATCAAAACAATTTAAAG 480
QY 1346 ATAAGAAGAAAAGATCAGAGAACCTTCAACAAATGAGGAGAAAGTAGATACAGATGATC 1405
DB 481 ATAAGAAGAAAAGATCAGAGAACCTTCAACAAATGAGGAGAAAGTAGATACAGATGATC 540
QY 1406 AGTTGTCCAAAGTATGAATCTCAACTTTCACAAATGAGGAGAAAGTAGATACAGATGATC 1465
DB 541 AGTTGTCCAAAGTATGAATCTCAACTTTCACAAATGAGGAGAAAGTAGATACAGATGATC 600
QY 1466 GAACCTGAAGGCTATTACGAGCAGACTCACAAGAGCCCTCCACCTTTTGTATCTCAGCAGC 1525
DB 601 GAACCTGAAGGCTATTACGAGCAGACTCACAAGAGCCCTCCACCTTTTGTATCTCAGCAGC 660
QY 1526 CTGCGAGTCTTGAAGAAGAGGTCATGATAGCTCATGCTCATCCACAGG--AAGTCTAC 1584
DB 661 CTGCGAGTCTTGAAGAAGAGGTCATGATAGCTCATGCTCATCCACAGCAAGTCTAC 720
QY 1585 AATGAATATGTACCC-----AGAGGCTGCAAGAAATAATGCCATTTCACAGATAC 1640
DB 721 AATGAATATGTAAACCCAGAGGGTCCCAAGAAATAATGCCATTTCACAGATAC 780
QY 1641 ACTCGGCCAGTCAG 1654
```


/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:5550711"
/clone_lib="NIH_MGC_67"
/tissue_type="Retinoblastoma"
/lab_host="DH10B (phage-resistant)"
/note="Organ: eye; vector: pCMV-SPOK6; Site_1: NotI;
Site_2: SalI; Cloned unidirectionally. Primer: Oligo dT.
Average insert size 1.75 kb. Library constructed by Life
Technologies."
BASE COUNT 292 a 255 c 266 g 322 t
ORIGIN

Query Match 28.7%; Score 652; DB 13; Length 1115;
Best Local Similarity 98.3%; Pred. No. 1-6e-135;
Matches 680; Conservative 0; Mismatches 10; Indels 2; Gaps 2;
QY 1464 TCCAACTGAAGCTATTACGACACACTCACAGAGCCCTCCCACTTTGATTCAGCA 1523
Db 2 TCTCAATGAAGGCTATTACGAGCAGACTCACAGAGCCCTCCCACTTTGATTCAGCA 61
QY 1524 GCCTGCAGCTTTCGAAGAAGAGGCTCATGATAGCTCATGCTCATCCACAGCAACTCTA 1583
Db 62 GCCTGCAGCTTTCGAAGAAGAGGCTCATGATAGCTCATGCTCATCCACAGGAATCTA 121
QY 1584 CAATGAATATGTACCCAGAGGGTGCAGGAATAAATGCCATTCCACATTCACATACACT 1643
Db 122 CAATGAATATGTACCCAGAGGGTGCAGGAATAAATGCCATTCCACATTCACATACACT 181
QY 1644 CCGCCAGTCAGAGGATCTATTACCAACCATCATGACTACCATCATATTCCTCATCATCA 1703
Db 182 CGCCAGTCAGAGGATCTATTACCAACCATCATGACTACCATCATATTCCTCATCATCA 241
QY 1704 CCACCAACCAACCAACCACTCTCACAGTCACAGCCAGCCCTACTCTCGGAGAGCTGAA 1763
Db 242 CCACCAACCAACCAACCACTCTCACAGTCACAGCCAGCCCTACTCTCGGAGAGCTGAA 301
QY 1764 AGATGCCGGCTGCGGCACCTTTGGCCGTGATGATGATGATGATGATGATGATGATGAT 1823
Db 302 AGATGCCGGCTGCGGCACCTTTGGCCGTGATGATGATGATGATGATGATGATGATGAT 361
QY 1824 CAGCGATGCGCTAGCAATTTGGTCTCTTTTACTGAAGGCTTATCACTGCTTTAAGTAC 1883
Db 362 CAGCGATGCGCTAGCAATTTGGTCTCTTTTACTGAAGGCTTATCACTGCTTTAAGTAC 421
QY 1884 TTCTGTTGCTGTTCTGTCATGCTGCTCATGAAATPAGTGACTTTGCTGCTTCTACT 1943
Db 422 TTCTGTTGCTGTTCTGTCATGCTGCTCATGAAATPAGTGACTTTGCTGCTTCTACT 481
QY 1944 AAAGCTTGGCATGACCGTTAAGCAGGCTGCTCTTTATATGATGATGATGATGATGATGAT 2003
Db 482 AAAGCTTGGCATGACCGTTAAGCAGGCTGCTCTTTATATGATGATGATGATGATGATGAT 541
QY 2004 GTATCTTGGCAATGGCAACAGGAATTTTCATGCTCATGCTGAAATGTTTCTATG 2063
Db 542 GTATCTTGGCAATGGCAACAGGAATTTTCATGCTCATGCTGAAATGTTTCTATG 601
QY 2064 GATATTTGCACTTACTGCTGCTTATTCATGATGATGCTGCTGCTGATATGCTA-CCTG 2122
Db 602 GATATTTGCACTTACTGCTGCTTATTCATGATGATGCTGCTGCTGATATGCTGACCTG 661
QY 2123 AAATGCTGCACAA-TGATGCTAGTCACCATGG 2153
Db 662 AAATGCTGCACAAATGATGCTAATGAACATGG 693

RESULT 6
LOCUS BM978669/c
DEFINITION BM978669 674 bp mRNA Linear EST 21-MAR-2002
ACCESSION UI-CF-DU1-ads-i-15-0-UI.s1 UI-CF-DU1 Homo sapiens cDNA clone
VERSION BM978669 UI-CF-DU1-ads-i-15-0-UI 3', mRNA sequence.
BM978669.1 GI:19598338

KEYWORDS
SOURCE
ORGANISM
REFERENCE
AUTHORS
TITLE
JOURNAL
MEDLINE
COMMENT

EST.
human.
Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 674)
Bonaldo,M.F., Lennon,G. and Soares,M.B.
Normalization and subtraction: two approaches to facilitate gene
discovery
Genome Res. 6 (9), 791-806 (1996)
9704477
Contact: McCray, PB
McCray Lab
University of Iowa
2024 University of Iowa Med Labs, Iowa City, IA 52242, USA
Tel: 319 356 4866
Fax: 319 356 7171
Email: paul-mccray@uiowa.edu
Tissue Procurement: Dr. M. J. Welsh, University of Iowa
CDNA Library preparation: Dr. M. Bento Soares, University of Iowa
CDNA Library Arrayed by: Dr. M. Bento Soares, University of Iowa
DNA Sequencing by: Dr. M. Bento Soares, University of Iowa
Clone Distribution: Researchers may obtain clones from Research
Genetics (www.resgen.com).
The following repetitive elements were found in this cDNA
sequence: 1-23, >AT-rich#Low_complexity (matched complement)
Seq primer: M13 FORWARD
POLYA-yes.

FEATURES
source

Location/Qualifiers
1..674
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="UI-CF-DU1-ads-i-15-0-UI"
/clone_lib="UI-CF-DU1"
/tissue_type="Primary Lung Epithelial Cells"
/dev_stage="Adult"
/lab_host="DH10B (Life Technologies) (T1 phage resistant)"
/note="Organ: Lung; Vector: pT73-Pac (Pharmacia) with a
modified polylinker; Site_1: EcoR 1; Site_2: Not I;
UI-CF-DU1 is a normalized cDNA library containing the
following tissue(s): Primary Lung Epithelial Cells The
library was constructed according to Bonaldo, Lennon and
Soares, Genome Research, 6:791-806, 1996. First strand
cDNA synthesis was primed with an oligo-dT primer
containing a Not I site. Double stranded cDNA was ligated
to an EcoR I adaptor, digested with Not I, and cloned
directionally into pT73-Pac vector. The oligonucleotide
used to prime the synthesis of first-strand cDNA contains
a library tag sequence that is located between the Not I
site and the (dT)18 tail. The sequence tag for this
library is GGCTGTAGGC.
TAG_LIB=UI-CF-DU1
TAG_TISSUE=Lung Epithelial Cells Tissue nos 359-368
TAG_SEQ=GGCTGTAGGC

BASE COUNT
ORIGIN

194 a 136 c 147 g 197 t
Query Match 28.7%; Score 649.8; DB 14; Length 674;
Best Local Similarity 98.8%; Pred. No. 4.6e-135;
Matches 663; Conservative 0; Mismatches 7; Indels 1; Gaps 1;

QY 687 GAGCCGGGTGAGCGGCTGGCTGTAGGAAACAAATGAATCTCAGGAGTGTTCATGATCA 745
Db 674 GAGCCGGGTGAGCGGCTGGCTGTAGGAAACAAATGAATCTCAGGAGTGTTCATGATCA 615
QY 746 AAGCTTTATGTATTCCAGAAACAAATGAAATCTCAGGAGTGTTCATGATCA 805
Db 614 AAGCTTTATGTATTCCAGAAACAAATGAAATCTCAGGAGTGTTCATGATCA 555
QY 806 AGCTTACTGACATCTCATGGCATGGCATCCAGGTTCCGCTGAATGCAACAGAGTCAACT 865
Db 554 AGCTTACTGACATCTCATGGCATGGCATCCAGGTTCCGCTGAATGCAACAGAGTCAACT 495

```
QY 866 ATCTCTGTCAGCCATCATCAACCAAAATTGATGCTAGATCTTGTCTGATTCATACAAAGTG 925
Db 494 ATCTCTGTCAGCCATCATCAACCAAAATTGATGCTAGATCTTGTCTGATTCATACAAAGTG 435
QY 926 AAAGAAGCGTGAATCCCTCCAAAGACCTATTTCATTAACAATAGCCCTGGGTTGGTGGTT 985
Db 434 AAAGAAGCGTGAATCCCTCCAAAGACCTATTTCATTAACAATAGCCCTGGGTTGGTGGTT 375
QY 986 TTATAGCCATTTCCATCATCAGTTTCCCTGCTCTGCTGGGGTTATCTTATGCTCTCA 1045
Db 374 TTATAGCCATTTCCATCATCAGTTTCCCTGCTCTGCTGGGGTTATCTTATGCTCTCA 315
QY 1046 TCAATCGGCTGTTTTTTCAAAATTTCTCCCTGAGTTTCTTGTGGCAGTGGCCCTTGGGACTT 1105
Db 314 TCAATCGGCTGTTTTTTCAAAATTTCTCCCTGAGTTTCTTGTGGCAGTGGCCCTTGGGACTT 255
QY 1106 TGAGTGGTGATGCTTTTTTACACCTTCTTCCACATTTCTCATGCAAGTCACCAACATAGTC 1165
Db 254 TGAGTGGTGATGCTTTTTTACACCTTCTTCCACATTTCTCATGCAAGTCACCAACATAGTC 195
QY 1166 ATAGCCATGAGAACCAGCAATGGAAATGAANAGAGGACCACTTTTTCAGTCTATCTCTTT 1225
Db 194 ATAGCCATGAGAACCAGCAATGGAAATGAANAGAGGACCACTTTTTCAGTCTATCTCTTT 135
QY 1226 CTCAAAACATAGAAGAAAGTGCTATTTCATTTCCACCTGGAAGGCTCTAACAGCTCTAG 1285
Db 134 CTCAAAACATAGAAGAAAGTGCTATTTCATTTCCACCTGGAAGGCTCTAACAGCTCTAG 75
QY 1286 GAGGCTCTGATTTTCATGTTTCTTGTGGAACATGTCCTCACATTCATGATCAACAATTTAAAG 1345
Db 74 GAGGCTCTGATTTTCATGTTTCTTGTGGAACATGTCCTCACATTCATGATCAACAATTTAAAG 15
QY 1346 ATAAAGAAAGAAAA 1358
Db 14 AAAAAAAGAAAAA 2

RESULT 7
BG168169
LOCUS 602341563P1 NIH_MGC_89 Homo sapiens cDNA clone IMAGE:449591 5',
DEFINITION mRNA sequence.
ACCESSION BG168169
VERSION BG168169.1 GI:12674872
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 957)
AUTHORS NIH-MGC http://mgi.nci.nih.gov/.
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL Unpublished (1999)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgapbs-r@mail.nih.gov
Tissue Procurement: ATCC
cDNA Library Preparation: Life Technologies, Inc.
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LIML)
DNA Sequencing by: Incyte Genomics, Inc.
Clone Distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LIML at:
http://image.llnl.gov
Plate: LAM10233 row: m column: 16
High quality sequence start: 14
High quality sequence stop: 659.
Location/Qualifiers
1. .957
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:449591"
/tissue_lib="NIH_MGC_89"
/tissue_type="hypernephroma, cell line"
/lab_host="DH10B (phage-resistant)"

FEATURES
source
```

```
/note="Organ: kidney; Vector: pCMV-SPORT6; Site_1: NotI;
Site_2: Sali; Cloned unidirectionally; oligo-dr' primed.
Average insert size 1.3 kb. Library enriched for
full-length clones and constructed by Life Technologies.
Note: this is a NIH_MGC Library."
BASE COUNT 321 a 229 c 204 g 203 t
ORIGIN
Query Match 27.7%; Score 627.6; DB 12; Length 957;
Best Local Similarity 89.6%; Pred. No. 4.6e-130;
Matches 731; Conservative 0; Mismatches 79; Indels 6; Gaps 5;
QY 288 CGACCATCATCTCAGACACGAGCATCAC-TCAGACCATGAGC-GTCACCTCAGACCATGAG 345
Db 1 CGACCATCATCTCAGACACGAGCATCACGTGACACCATGAGCGGTGCTCCTCAGACATGAG 60
QY 346 C-ATCCTCAGACACGAGCATCACCTCTGACCATGATCATCTCTCACCATAATCATGC 404
Db 61 CAATCCTCAGACACGAGCATCACCTCTGACCATGATCATCTCTCACCATAATCATGC 120
QY 405 TGC-TTCTGCTAAATAAGCGAAAGAGCTCTTTGGCCAGACCATGACTCAGATAGTTTCA 463
Db 121 TGCCTTCTGCTAAATAAGCGAAAGAGCTCTTTGGCCAGACCATGACTCAGATAGTTTCA 180
QY 464 GTAAAGATCTCTAGAAACAGCGAGGGAAGAGCTCACCAGCAGAACATGCCAGTGGTA 523
Db 181 GTAAAGATCTCTAGAAACAGCGAGGGAAGAGCTCACCAGCAGAACATGCCAGTGGTA 240
QY 524 GAAGGAATGTCAAGGACAGTGTGTAGTGTAGTGAAGTGAAGTCAACTGTGTACACACTG 583
Db 241 GAAGGAATGTCAAGGACAGTGTGTAGTGTAGTGAAGTGAAGTCAACTGTGTACACACTG 300
QY 584 TCTCTGAAGGAACCTCATTCTTAGAGACATAGAGACTCCAAGACCTGGAAACTCTTCC 643
Db 301 TCTCTGAAGGAACCTCATTCTTAGAGACATAGAGACTCCAAGACCTGGAAACTCTTCC 360
QY 644 CCAAGATGTAAAGCAGCTCCACCTCCACCTGTCACATCAAGAGCGGGTGAGCCGGC 703
Db 361 CCAAGATGTAAAGCAGCTCCACCTCCACCTGTCACATCAAGAGCGGGTGAGCCGGC 420
QY 704 TGCTGGTAGGAAACAAATGAATCTGTGAGTGAGCCGCCGAAAGGCTTTATGTATTCCA 763
Db 421 TGCTGGTAGGAAACAAATGAATCTGTGAGTGAGCCGCCGAAAGGCTTTATGTATTCCA 480
QY 764 GAAACACAATGAAATCTCTCAGGAGTGTTCATGATCAATCAAGAGCTTACTGACATCTCATG 823
Db 481 GAAACACAATGAAATCTCTCAGGAGTGTTCATGATCAATCAAGAGCTTACTGACATCTCATG 540
QY 824 GCATGGCATCCAGGTTCCGCTGAATGCAACAGAGTTCAACTATCTCTGTCCAGCCATCA 883
Db 541 GCATGGCATCCAGGTTCCGCTGAATGCAACAGAGTTCAACTATCTCTGTCCAGCCATCA 600
QY 884 TCAACCAATTTGATGCTAGATCTTGTGATTCATACAAAGTGAAGAGGCTGAAATCC 943
Db 601 TCAACCAATTTGATGCTAGATCTTGTGATTCATACAAAGTGAAGAGGCTGAAATCC 660
QY 944 CTCCAAAGACCTATTTCATTAACAATAGCCCTGGGTTGGTGGTGTATATAGCAATTTCCATCA 1003
Db 661 TTCCAAAGAGCTAATTCATTTACAAATAGGCTGAGTTGGGGTATATAGCAATTTCCATCA 720
QY 1004 TCAGTTTCCCTGCTCTGCTGGGGTTATCTTACTGCTCTCTCAATGAATCGGGTGTGTTTCA 1063
Db 721 ATCAGTACCTG--CACTGGCGGGGTAAACTAAACGGCCCTTAATGATCGGGGTGCCACA 778
QY 1064 AATTTCTCTGAGTGTTCCTTGTGGCAGTGGCCGCTTG 1099
Db 779 AATCCCTGGCATTCTCTGTGTGACTGGCTGGCAGCTTG 814

RESULT 8
BG530601
LOCUS 602560191F1 NIH_MGC_61 Homo sapiens cDNA clone IMAGE:469787 5',
DEFINITION
```



```

mRNA sequence.
BG530601
BG530601.1 GI:13522138
EST.
SOURCE human.
ORGANISM Homo sapiens
REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
AUTHORS Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
TITLE NIH-MGC http://mgi.nci.nih.gov/.
JOURNAL National Institutes of Health, Mammalian Gene Collection (MGC)
COMMENT Unpublished (1999)
Contact: Robert Strausberg, Ph.D.
Email: rstraus@nih.gov
Tissue Procurement: ATCC
CDNA Library Preparation: CLONTECH Laboratories, Inc.
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: InCyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
Plate: LCM1527 row: f column: 23
High quality sequence stop: 623.
Location/Qualifiers
1..746
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:4697878"
/clone_lib="NIH_MGC_61"
/tissue_type="embryonal carcinoma"
/lab_host="DH10B (T1 phage-resistant)"
/notes="Organ: testis; Vector: pONR-LiH (Clontech); Site_1:
SfiI (ggccgctcgccc); Site_2: SfiI (ggcattatgcc);
Double-stranded cDNA was prepared from cell line RNA. 5'
and 3' adaptors were used in cloning as follows: 5'
adaptor sequence: 5'-CACGCCATATGCCC-3' and 3' adaptor
sequence: 5'-ATTCTAGAGCGCGCGGCACATG-DT(30)BN-3'
(where B = A, C, G or N = A, C, G, or T). Average
insert size 1.75 kb (range 0.9-4.0 kb). 15/15 colonies
contained inserts by PCR. This library was enriched for
full-length clones and was constructed by Clontech
Laboratories (Palo Alto, CA). Note: this is a NIH_MGC
Library."
BASE COUNT 195 a 151 c 179 g 221 t
ORIGIN

Query Match 26.5%; Score 600.2; DB 12; Length 746;
Host local Similarity 96.4%; Pred. No. 6.1e-124;
Matches 647; Conservative 0; Mismatches 18; Indels 6; Gaps 3;

QY 1575 GGAAGTCTACAAATATGTACCCAGGCTGCAAGAAATAAATGCCATTTCACATTCCA 1614
Db 1 GGAAGTCTACAAATATGTACCCAGGCTGCAAGAAATAAATGCCATTTCACATTCCA 60

QY 1615 CGATACACTCGGCCAGTCAGACGATCTCATTACACCATCATCATCATCATTTCT 1694
Db 61 CGATACACTCGGCCAGTCAGACGATCTCATTACACCATCATCATCATCATTTCT 120

QY 1695 CCATCATCACCAACC-ACCAAAACACCATCCTCACATCATCACGCGCTACTCTCGGG 1753
Db 121 CCATCATCACCAACCAGCAAAACACCATCCTCACATCATCACGCGCTACTCTCGGG 180

QY 1754 AGGAGCTGAAAGATGCCGGCTGCCCACTTTGGCTGGATGGTCATATGCGTGGCC 1813
Db 181 AGGAGCTGAAAGATGCCGGCTGCCCACTTTGGCTGGATGGTCATATGCGTGGCC 240

QY 1814 TGCACAAATTCAGCGCATGGCTAGCAATATGGTGTGCTTTTACTGAAGGCTTATCAAGTG 1873
Db 241 TGCACAAATTCAGCGCATGGCTAGCAATATGGTGTGCTTTTACTGAAGGCTTATCAAGTG 300

QY 1874 GTTTTAAGTACTTCTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 1933
Db 301 GTTTTAAGTACTTCTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 360

```

```

QY 1934 CTGTTCTACTAAAGGCTJGCATGACCGTTTAAAGCAGGCTGCTCTTTTATAATGCTATGTCAG 1993
Db 361 CTGTTCTACTAAAGGCTJGCATGACCGTTTAAAGCAGGCTGCTCTTTTATAATGCTATGTCAG 420

QY 1994 CCATGCTGGCGTATCTTGGAAATGGCAACAGGAATTTTCATTTGGTTCATTTATGCTGAAAATG 2053
Db 421 CCATGCTGGCGTATCTTGGAAATGGCAACAGGAATTTTCATTTGGTTCATTTATGCTGAAAATG 480

QY 2054 TTTCTATGTTGGATATTTGCACCTTACTGCTGGCTTATTCATGATGCTGCTGCTGTTGATA 2113
Db 481 TTTCTATGTTGGATATTTGCACCTTACTGCTGGCTTATTCATGATGCTGCTGCTGTTGATA 540

QY 2114 TGTACTCTGAAATGCTGCACAATGATGCTAGTGACCATGGATGTAGCCGCTGGGGTATT 2173
Db 541 TGTACTCTGAAATGCTGCACAATGATGCTAGTGACCATGGATGTAGCCGCTGGGGTATT 599

QY 2174 TCTTTTACAGAATGCTGGGATGCTTTTGGGGTATTTTGGAAATTTATTTTCCATAT 2233
Db 600 ACTTTTACAGAATGCTGGGATGCTTTGGGGATTT---GAATTATGTACTGATATCCATAT 655

QY 2234 TTGAACATATA 2244
Db 656 TGGACCTAAAA 666

RESULT 9
BQ581865
LOCUS BQ581865 604 bp mRNA linear EST 20-JUN-2002
DEFINITION i110b02.y1 Human insulinoma Homo sapiens cDNA clone IMAGE:6029499
5' similar to TR:Q13433 Q13433 ESTROGEN REGULATED LIV-1 PROTEIN. ;,
mRNA sequence.
ACCESSION BQ581865
VERSION BQ581865.1 GI:21494761
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 604)
AUTHORS Melton,D., Brown,J., Kenty,G., Permutt,A., Lee,C., Kaestner,K.,
Lemishka,I., Searce,M., Brestelli,J., Gradwohl,G., Clifton,S.,
Hillier,L., Marra,M., Pape,D., Wylie,T., Martin,J., Blistain,A.,
Schmitt,A., Theising,B., Ritter,E., Ronko,I., Bennett,J., Cardenas
,M., Gibbons,M., McCann,R., Cole,K., Tsagareishvili,R., Williams,T.,
Jackson,Y. and Bowers,Y.
Endocrine Pancreas Consortium
Unpublished (2000)
Contact: Douglas Melton, Klaus H. Kaestner, & Hiroshi Inoue
Endocrine Pancreas Consortium
Harvard University, Howard Hughes Medical Institute
Dept of Molecular and Cellular Biology, 7 Divinity Ave, Cambridge,
MA 02138
Tel: 617-495-1812
Fax: 617-495-8557
Email: dmelton@biohp.harvard.edu
Library was constructed by Dr. J. J. Ferrer In vivo mass-excised to
pBluescript SK- by Dr. H. Inoue DNA sequencing by: Washington
University Genome Sequencing Center For information on obtaining a
clone please contact: Dr. Hiroshi Inoue (hinoue@im.wustl.edu)
Seq primer: -40RP from Gibco
High quality sequence stop: 431.
FEATURES
source
1..604
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:6029499"
/clone_lib="Human insulinoma"
/tissue_type="insulinoma"
/lab_host="DH10B (phage-resistant)"
/notes="Organ: pancreas; Vector: pBluescript SK-; Site_1:
XhoI; Site_2: EcoRI; Constructed with lambda ZAP11 system
(Stratagene) by Dr. J. Ferrer, in vivo mass-excised to

```

pBuescript SK- by Dr. H. Inoue following the Washington University protocol (http://genome.wustl.edu/est/lambda_protocol.shtml). Please contact Hiroshi Inoue, MD/PhD for further information on this library (Metabolism Division, Permut Laboratory, Washington University School of Medicine, Box 8127, 660 S Euclid Ave, St. Louis, MO 63110). Note: this is a Washington University Pancreas EST project library. "

BASE COUNT	163 a	143 c	130 g	168 t
ORIGIN				
Query Match	26.3%	Score 597.6;	DB 14;	Length 604;
best local Similarity	99.3%			
Matches 600;	Conservative	0;	Mismatches 4;	Indels 0;
		Gaps 0;		

[illegible]

REFSeq ID	BM948026	766 bp	linear	EST 14-MAR-2002
LOCUS	U1-M-EG0p-bvd-b-19-0-U1.r1	NIH_BMAP_EG0p	Mus musculus cDNA clone	
DEFINITION	IMAGE:5690418.5, mRNA sequence.			
ACCESSION	BM948026			
VERSION	BM948026.1	GI:19431616		
KEYWORDS	EST.			
SOURCE	house mouse.			
ORGANISM	Mus musculus			
	Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;			
	Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.			
REFERENCE	1 (bases 1 to 766)			

AUTHORS NIH-MGC <http://mgc.nci.nih.gov/>.
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL Unpublished (1999)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgapbs-r@mail.nih.gov
Tissue Procurement: Dr. James Lin, University of Iowa
cDNA Library preparation: Dr. M. Bento Soares, University of Iowa
cDNA Library Arrayed by: Dr. M. Bento Soares, University of Iowa
DNA Sequencing by: Dr. M. Bento Soares, University of Iowa
Clone Distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
<http://image.llnl.gov>
This clone was contributed by the Brain Molecular Anatomy Project
(BNAP)
The following repetitive elements were found in this cDNA
sequence: 318-399, >(TGG)n#Simple repeat (matched complement).

seq primer: p18.5.

FEATURES	Location/Qualifiers
source	1..766

Source

1. 766
/organism="Mus musculus"
/strain="C57BL/6"
/db_xref="taxon:10090"
/clone="IMAGE:5690418"
/clone_lib="NIH_BMAP_EG0p"
/tissue_type="whole brain"
/dev_stage="embryo 18.5 dpc"
/lab_host="DH10B (TI phase resistant)"
/note="Organ: brain; Vector: pYX-Asc; Site:1: Ecor I;
Site:2: Not I; The library was constructed according to
Bonaldo, Lennon and Soares, Genome Research, 6:791-806,
1995. Denatured mRNA was size fractionated on a 1% agarose
gel. First strand cDNA synthesis was primed with an
oligo-dT primer containing a Not I site. Double stranded
cDNA was size selected according to mRNA size fraction,
ligated with Ecor I adaptor, digested with Not I, and then
cloned directionally into pYX-Asc vector. The library tag
sequence located between the Not I site and the polyA tail
, is CAGCCACACAC. This library was created for the
University of Iowa Mouse Brain Molecular Anatomy Project
(BMAP): "Gene Discovery in the Developing Mouse Nervous
System", supported by National Institute of Mental Health
(NIMH), Hemin Chin, Ph.D., program coordinator."

201 a 208 c 192 g 165 t

BASE COUNT
BRTCN

Query Match 25.9%; Score 586.8; DB 14; Length 766;
Best Local Similarity 85.4%; Pred. No. 6.2e-121;
Matches 654; Conservative 0; Mismatches 112; Indels 0; Gaps 0;

QY	1348	AGAAGAAAAGAAATCAGAGAAACCTGAAATGATGATGATGGAGATTTAAGAACGAC	1407
Db	1	AAGAAAAAGGAAGATCAAGAAAAACCTGAAATGATGAGGATGGAGAGCAAGAACGAC	60
QY	1408	TTGTCCAAGTATGATCTCAACTTTCAACAAATCAGAGAAAAGTAGATACAGATGATCGA	1467
Db	61	CTGTCCAAATACGACTCTCAGCTTTCCTCAATGAAGAAAGGTGGACCCAGGGGAACGA	120
QY	1468	ACTGAAGGCTATTTACGAGCAGACTCACAGAGGCCCTCCACATTTGATTTCTCAGCAGCCT	1527
Db	121	CCTGAAGGCTATCTCGGAGCGGACTCCAGAGGCCCTCCCCCTTTGATTTCCACGACGGC	180
QY	1528	GCAGTCTTTGGAAAGAAGAGGTCATGATAGCTCATGCTCATCCACAGGAAGTCTACAAT	1587
Db	181	ACGATCTTTGGAAGGAAGAGGTCATGATAGCCCATGCCACCCACACAGAAGTCTACAAT	240
QY	1588	GAATATGTACCCAGAGGCTGCAAGAATAAATGCCATTTACATTTCCACGATACACTCGGC	1647
Db	241	GAATATGTGCCAGGGGCTGCAAGAACAAGTGCCATTTACACTTCCACGATACGCTGGGC	300
QY	1648	CAGTCAGACGATCTCATTTACACCGCATCATGACTACCATCATATTTCTCCATCATCACCA	1707
Db	301	CAGTCGAGGAGCTCATCCACCGCATCAGAGCTACCATTCACATTTCTGGACCAACCAAC	360

MHLINE:						
COMMENT						
20202663 Contact: Simpson A.J.G. Laboratory of Cancer Genetics Ludwig Institute for Cancer Research Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP, Brazil Tel: +55-11-2704922 Fax: +55-11-2707001 Email: asimpson@ludwig.org.br This sequence was derived from the FAPESP/LICK Human Cancer Genome Project. This entry can be seen in the following URL: (http://www.ludwig.org.br/seq/gethtml.pl?l1=R:&l2=KC-BT133-076.html&f3=lyd399&t4=1) Seq primer: puc 18 forward. Location/Qualifiers 1..743 /organism="Homo sapiens" /db_xref="Laxon:9606" /cclone_lib="BT133" /sex="female" /dev_stage="Adult" /note="Organ: breast; Vector: puc18; Site_1: Small; Site_2: A mini-library was made by cloning products derived from ORESTES PCR (O.S. Letters Patent application No. 196 716 - Ludwig Institute for Cancer Research) profiles into the pUC 18 vector. Reverse transcription of tissue mRNA and cDNA amplification were performed under low stringency conditions." BASE COUNT 178 a 155 c 175 g 227 t 8 others ORIGIN						
Query Match 25.2%; Score 571.4; DB 9; Length 743;						
Best local similarity 95.9%; Pred. No. 1.7e-117;						
Matches 648; Conservative 0; Mismatches 21; Indels 7; Gaps 6;						
QY	1595	TACCCAGAGGGTGCGAAGAATAATGCCATTACCATTTCCAGGATACACTCGGCACGTCTCAG	1654			
Dd	14	TACCCAGAGGGTGCGAAGAATAGATGCCATTACCATTTCCAGATFACACTCGGCACGTCA	73			
QY	1655	ACUATCTCATTCACCACCATCATGACTACCATCATATTCTCCATCATCACACACACCAA	1714			
Dd	74	ACUTCTCATTCACCACCATCATGACTACCATCATATTCTCCATCATCACACACACCAA	133			
QY	1715	ACCACCATCTTCACAGTCACAGCCCAGCGCTACTCTCGGGAGGAGCTGAAGATGCCGGG	1774			
Dd	134	ACCACCATCTTCACAGTCACAGCCCAGCGCTACTCTCGGGAGGAGCTGAAGATGCCGGG	193			
QY	1775	TGGCCTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG	1834			
Dd	194	TGUCCACATCTGGCGTGGATGGTGAATGGTGGATGGCGCTGCACAATTCAGCGATGGCG	253			
QY	1835	TAGTAATTCGTCGCTTTTTACTGAAGCGTTATCAAGTGTGTTAAAGTACTTCTGTTGCTG	1894			
Dd	254	TAGCAATTCGTCGCTTTTTACTGAAGCGTTATCAAGTGTGTTAAAGTACTTCTGTTGCTG	313			
QY	1895	TGTTCTCTCATGAGTTGCCTCATGAATTTAGTGACTTTGGCTGTCTACTAAAGCTTGCCA	1954			
Dd	314	TGTTCTCTCATGAGTTGCCTCATGAATTTAGTGACTTTGGCTGTCTACTAAAGCTTGCCA	373			
QY	1955	TGACCGTTTAAGCAGGCTGTCTTTTAATATGATTTGTGACGCGATGCT-GGCGTATCTTGA	2013			
Dd	374	TGACCGTTTAAGCAGGCTGTCTTTTAATATGATTTGTGACGCGATGCTGGCGTATCTTGA	433			
QY	2014	ATGGCAACAGGAATTTTCATTGGTCATTTATCGTAAAATGTTTCTATGTGGATATTTCGA	2073			
Dd	434	ATGGCAACAGGAATTTTCATTGGTCATTTATCGTAAAATGTTTCTATGTGGATATTTCGA	493			
QY	2074	CTTTACCTGCTGGCTTATTTATGATGATGCTGCTGTGATATGGTACCTGAATATCTGAC	2133			
Dd	494	CTTTACCTGCTGGCTTATTTATGATGATGCTGCTGTGATATGGTACCTGAATATCTGAC	553			
QY	2134	AATGATG-CTAGTGACCCAT-GGATGTAGCCG-CTGGGGGTATTTTCTTTTTC-TAG-A	2189			
Dd		AAAAATGATGCTAGTGACCCATGGATGTAGCCGCTGGGGGTATTTTCTTTTTCAGAAAT				

Db	554	AATGATGCCTAGTGACCATGGGATGTATGCCCTGGGGGGTATTTCTTTTACAAAAATGC	613
Qy	2190	TGGGATGCTTTTGGG--TTTGTGGAATATGTTACTTATTTCCATATTTGAACATATAAATC	2247
Db	614	TGGGATGCTTTTGGGTTTGGGAATTTATGTTACTTATTTTCATATTTGACATAAANCC	673
Qy	2248	GTGTTTTCGTATATAAATTT	2263
Db	674	GGGTTTCGATNAATTT	689
RESULT	13		
AU079709			
LOCUS			
DEFINITION	AU079709 Sugano mouse brain mncb Mus musculus cDNA clone	829 bp	linear EST 12-JUL-2000
	5' mRNA sequence.		
ACCESSION	AU079709		
VERSION	AU079709.1	GI:6084464	
KEYWORDS	EST.		
SOURCE	house mouse.		
ORGANISM	Mus musculus		
	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;		
REFERENCE	Mammalia; Eutheria; Rodentia; Sclurognathi; Muridae; Murinae; Mus.		
AUTHORS	1 (bases 1 to 829)		
	Hashimoto, K., Kusuda, J., Tanuma, R., Ito, A., Hirata, M., Toyoda, A.,		
	Suzuki, Y., Sasaki, M. and Sugano, S.		
TITLE	Isolation of full-length cDNA clones from a mouse brain cDNA		
	library made by oligo-capping method		
JOURNAL	Unpublished (1999)		
COMMENT	Contact: Katsuyuki Hashimoto		
	Division of Genetic Resources		
	National Institute of Infectious Diseases		
	23-1, Toyama 1-chome, Shinjuku-ku, Tokyo 162-8640, Japan		
	Email: khash@nih.go.jp		
	URL: http://www.nih.go.jp/yoken/genbank/.		
FEATURES	Location/Qualifiers		
source	1..829		
	/organism="Mus musculus"		
	/strain="C57BL"		
	/db_xref="taxon:10090"		
	/clone_lib="MNCB-4665"		
	/clone_lib="Sugano mouse brain mncb"		
	/sex="female"		
	/dev_stage="adult"		
	/lab_host="TOP10"		
	/note="Organ: brain; Vector: pME18S-FL3; 1st strand cDNA		
	was primed with an oligo(dT) primer		
	ATGGGCTTTTTTTTTTTT; double-stranded cDNA was		
	ligated to a DraIII adaptor [TCTTGGCTACTGG], digested and		
	cloned into distinct DraIII sites of the pME18S-FL3. XhoI		
	sites just outside the DraIII sites can be used to isolate		
	the cDNA insert. Size selection was performed to exclude		
	fragments <1.5 kb. Library was constructed by Sugano et		
	al.(University of Tokyo, Institute of Medical Science).		
	Custom primer used for sequencing: 5' end primer		
	[CTTCTGCTCTAAAGCTGG], 3' end primer		
	[CGACCTGCAGCTCGACACA].		
BASE COUNT	231 a 220 c 188 g 187 t	3	others
ORIGIN			
Query Match	25.0%	Score 567.4;	DB 9; Length 829;
Best Local Similarity	84.5%;	Pred. No. 1.4e-116;	
Matches	695;	Conservative 0; Mismatches 119; Indels 8; Gaps 5;	
Qy	893	TTGATGCTAGATCTTGTCGTATTCATA----	CAAGTGAAGAAGGCGTGAAATCCCTCCAA 949
Db	1	TTGATGCTCGGCTTGTCGTATTCATACAGCAAGTGAGAGAAGCGAAGATCCCTCCAA	60
Qy	950	AGACCTATTATTACAATACCGTTGGTGGTTTATAGACCATTTCCATCATCAGTT	1009
Db	61	AGACCTATTCTTTACAATAGCGTGGCTTGCGTCTCATAGCCATTTCCATCATCAGTT	120
Qy	1010	TCCTGTCTCTCTCGGGGCTTATCTTACTTGCCCTCTCATCAATCGGCTGTTTTCAAATTC	1069

GenCore version 5.1.1.3
Copyright (c) 1993 - 2002 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: November 28, 2002, 04:15:57 ; Search time 143 seconds

(without alignments)

6107.985 Million cell updates/sec

Title: us-09-642-034-4

Perfect score: 2268

Sequence: 1 atggcgaggagattatctgt.....tggttcgtataaatlctag 2268

Scoring table: IDENTITY_NUC

Gapop 10.0 , Gapext 1.0

Searched: 341543 seqs, 192557720 residues

Total number of hits satisfying chosen parameters: 683086

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Published_Applications_NA.*

1: /cgn2_6/ptodata/1/pubpna/us07_PUBCOMB.seq.*
2: /cgn2_6/ptodata/1/pubpna/PCT_NEW_PUB.seq.*
3: /cgn2_6/ptodata/1/pubpna/US06_NEW_PUB.seq.*
4: /cgn2_6/ptodata/1/pubpna/US06_PUBCOMB.seq.*
5: /cgn2_6/ptodata/1/pubpna/US07_NEW_PUB.seq.*
6: /cgn2_6/ptodata/1/pubpna/PCTUS_PUBCOMB.seq.*
7: /cgn2_6/ptodata/1/pubpna/US08_NEW_PUB.seq.*
8: /cgn2_6/ptodata/1/pubpna/US08_PUBCOMB.seq.*
9: /cgn2_6/ptodata/1/pubpna/US09_NEW_PUB.seq.*
10: /cgn2_6/ptodata/1/pubpna/US09_PUBCOMB.seq.*
11: /cgn2_6/ptodata/1/pubpna/US10_NEW_PUB.seq.*
12: /cgn2_6/ptodata/1/pubpna/US10_PUBCOMB.seq.*
13: /cgn2_6/ptodata/1/pubpna/US60_NEW_PUB.seq.*
14: /cgn2_6/ptodata/1/pubpna/US60_PUBCOMB.seq.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1141.4	50.3	1193	10	US-09-925-300-591
2	319.2	14.1	337	9	US-09-920-455-250
3	319.2	14.1	5330	10	US-09-789-561-14
4	127.6	5.6	4573	10	US-09-880-107-1649
5	62.2	2.7	574	10	US-09-864-761-228
6	62.2	2.7	669	10	US-09-864-761-17051
7	59.8	2.6	285	9	US-09-286-488-64
8	59.8	2.6	285	10	US-09-737-178-64
9	59.8	2.6	342	9	US-09-286-488-65
10	59.8	2.6	342	10	US-09-737-178-65
11	59.8	2.6	351	9	US-09-286-488-60
12	59.8	2.6	351	10	US-09-737-178-60
13	59.8	2.6	356	9	US-09-286-488-63
14	59.8	2.6	356	10	US-09-737-178-63
15	59.8	2.6	358	9	US-09-286-488-58
16	59.8	2.6	358	10	US-09-737-178-58
17	59.8	2.6	363	9	US-09-286-488-66
18	59.8	2.6	363	9	US-09-286-488-67
19	59.8	2.6	363	10	US-09-737-178-66

c 20	59.8	2.6	363	10	US-09-737-178-67	Sequence 67, Appl
c 21	59.8	2.6	409	9	US-09-286-488-59	Sequence 59, Appl
c 22	59.8	2.6	409	10	US-09-737-178-59	Sequence 59, Appl
c 23	59.8	2.6	410	9	US-09-286-488-61	Sequence 61, Appl
c 24	59.8	2.6	410	10	US-09-737-178-61	Sequence 61, Appl
c 25	59.8	2.6	416	9	US-09-286-488-62	Sequence 62, Appl
c 26	59.8	2.6	416	10	US-09-737-178-62	Sequence 62, Appl
c 27	59.8	2.6	1271	9	US-09-286-488-5	Sequence 5, Appl
c 28	59.8	2.6	1271	9	US-09-286-488-5	Sequence 42, Appl
c 29	59.8	2.6	1271	10	US-09-737-178-5	Sequence 5, Appl
c 30	59.8	2.6	1271	10	US-09-737-178-5	Sequence 42, Appl
c 31	59.8	2.6	1821	9	US-09-286-488-6	Sequence 6, Appl
c 32	59.8	2.6	1821	10	US-09-737-178-6	Sequence 6, Appl
c 33	59.4	2.6	1635	10	US-09-864-761-20241	Sequence 20241, A
c 34	59.4	2.6	1973	10	US-09-864-761-3471	Sequence 3471, A
c 35	58.8	2.6	4223	9	US-09-286-488-7	Sequence 7, Appl
c 36	58.8	2.6	4223	9	US-09-286-488-45	Sequence 45, Appl
c 37	58.8	2.6	4223	10	US-09-737-178-7	Sequence 7, Appl
c 38	58.8	2.6	4223	10	US-09-737-178-45	Sequence 45, Appl
c 39	57.8	2.5	468	10	US-09-864-761-2321	Sequence 2321, A
c 40	57.2	2.5	18596	9	US-09-954-531-124	Sequence 124, App
c 41	57.2	2.5	18596	9	US-09-954-531-348	Sequence 348, App
c 42	57.2	2.5	18596	10	US-09-880-107-1590	Sequence 1590, Ap
c 43	57.2	2.5	18596	10	US-09-967-768A-119	Sequence 119, App
c 44	56.8	2.5	464	10	US-09-864-761-764	Sequence 764, App
c 45	56.8	2.5	583	10	US-09-864-761-20772	Sequence 20772, A

ALIGNMENTS

RESULT 1

US-09-925-300-591
; Sequence 591, Application us/09925300
; Patent No. US20020151681A1
; GENERAL INFORMATION:
; APPLICANT: Craig Rosen,
; APPLICANT: Steve Ruben,
; TITLE OF INVENTION: Nucleic Acids, Proteins and Antibodies
; FILE REFERENCE: PA101
; CURRENT APPLICATION NUMBER: US/09/925,300
; CURRENT FILING DATE: 2001-08-10
; PRIOR APPLICATION NUMBER: PCT/US00/05988
; PRIOR FILING DATE: 2000-03-08
; PRIOR APPLICATION NUMBER: 60/124,270
; PRIOR FILING DATE: 1999-03-12
; NUMBER OF SEQ ID NOS: 1890
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 591
; LENGTH: 1193
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-925-300-591

Query Match 50.3%; Score 1141.4; DB 10; Length 1193;
Best Local Similarity 99.7%; Pred. No. 3.6e-308;
Matches 1139; Conservative 4; Mismatches 0; Indels 0; Gaps 0;

QY	539	ACAGTGTCTAGTCTAGTGAAGTGACCTCAACTGTGTACAACTACTGTCTCTGAAGGAATCTC	598
DB	1	ACAGTGTCTAGTCTAGTGAAGTGACCTCAACTGTGTACAACTACTGTCTCTGAAGGAATCTC	60
QY	599	ACTTTCTAGAGACAATAGAGACTCCAAAGACCTCGGAGCCGGTGGCTGTAGTAAAGCA	658
DB	61	ACTTTCTAGAGACAATAGAGACTCCAAAGACCTCGGAGCCGGTGGCTGTAGTAAAGCA	120
QY	659	GCTCCACTCCAGCCAGTGTACATCAAAAGAGCCGGTGGCTGTAGTAAAGCA	718
DB	121	GCTCCACTCCAGCCAGTGTACATCAAAAGAGCCGGTGGCTGTAGTAAAGCA	180
QY	719	CAATCAATCTGTGAGTGGAGCCCGGAGGCTTTATGTTATTTCCAGAAACACAAATGAAA	778
DB	181	CAATCAATCTGTGAGTGGAGCCCGGAGGCTTTATGTTATTTCCAGAAACACAAATGAAA	240

QY 779 ATCCCTCAGGAGTGTCTTCAATGATCAAAAGCTACTCAGATCTCATGCGCATGGCATCCAGG 838
Db 241 ATCCCTCAGGAGTGTCTTCAATGATCAAAAGCTACTCAGATCTCATGCGCATGGCATCCAGG 300
QY 839 TTCCGGTGAATGCACAGAGTTCAACTATCTCTGTCCAGCATCATCAACCAAAATTTGATG 898
Db 301 TTCCGGTGAATGCACAGAGTTCAACTATCTCTGTCCAGCATCATCAACCAAAATTTGATG 360
QY 899 CTGATGCTTCTGCTGATTCATCAAACTGAAAGAGGCTGAATCCCTCCAAAGACCTATT 958
Db 361 CTGATGCTTCTGCTGATTCATCAAACTGAAAGAGGCTGAATCCCTCCAAAGACCTATT 420
QY 959 CATTACAAATAGCCCTGGGTGGTGTATATAGCCATTTCATCATCAGTTTCCCTGTCTC 1018
Db 421 CATTACAAATAGCCCTGGGTGGTGTATATAGCCATTTCATCATCAGTTTCCCTGTCTC 480
QY 1019 TGCCTGGGGTTATCTTTAGTGGCTCTCATGATCTGGGTGTTTTTCAAAATTTCTGCTGATTT 1078
Db 481 TGCCTGGGGTTATCTTTAGTGGCTCTCATGATCTGGGTGTTTTTCAAAATTTCTGCTGATTT 540
QY 1079 TCTTCTGCGCACTGGCCGTTGGGACTTTCAGTGGTGATGCTTTTACACCTTCTTCCAC 1138
Db 541 TCTTCTGCGCACTGGCCGTTGGGACTTTCAGTGGTGATGCTTTTACACCTTCTTCCAC 600
QY 1139 ATTCTCATGCAAGTCACCACTAGTCTATAGCCATGAAGAACCCAGCAATGGAATGAATA 1198
Db 601 ATTCTCATGCAAGTCACCACTAGTCTATAGCCATGAAGAACCCAGCAATGGAATGAATA 660
QY 1199 GAGGAGCCACTTTTTCAGTCTATCTGCTTCTCAAAACATAGAAAGTGCCTATTCTTGAAT 1258
Db 661 GAGGAGCCACTTTTTCAGTCTATCTGCTTCTCAAAACATAGAAAGTGCCTATTCTTGAAT 720
QY 1259 CCACCTGCAAGGCTCTAACAGCTCTAGGAGGCTCTATTTTTCATGCTTCTCTCAACATG 1318
Db 721 CCACCTGCAAGGCTCTAACAGCTCTAGGAGGCTCTATTTTTCATGCTTCTCTCAACATG 780
QY 1319 TCTTCTACATTTGATCAAACTTTTAAAGATAAGAAAGAAAGAAATCAGAAAGCTGAAA 1378
Db 781 TCTTCTACATTTGATCAAACTTTTAAAGATAAGAAAGAAAGAAATCAGAAAGCTGAAA 840
QY 1379 ATGATCATGATGCTGAGATTAAGAGAGGATTTGTCCAAGTATGAATCTCAACTTTTCAACA 1438
Db 841 ATGATCATGATGCTGAGATTAAGAGAGGATTTGTCCAAGTATGAATCTCAACTTTTCAACA 900
QY 1439 ATGAGCAGAAAGCTAGATACAGATGATCAACTGAAGGCTATTTACGAGCAGACTCACAG 1498
Db 901 ATGAGCAGAAAGTATGATACAGATGATCAACTGAAGGCTATTTACGAGCAGACTCACAG 960
QY 1499 AGCCCTTCCACCTTTTCACTTCTCAGCAGCCTGCGAGTCTTGGAAAGAGAGAGTCAATGATAG 1558
Db 961 AGCCCTTCCACCTTTTCACTTCTCAGCAGCCTGCGAGTCTTGGAAAGAGAGAGTCAATGATAG 1020
QY 1559 CTCATGCTCATCCACAGGAAGTCTACATGAATATGTACCCAGAGGGTGGCAAGAAATAAT 1618
Db 1021 CTCATGCTCATCCACAGGAAGTCTACATGAATATGTACCCAGAGGGTGGCAAGAAATAAT 1080
QY 1619 GCCATTTCATATTTCCACCATACACTTCGGCCAGTCTCAGAGCATCTCATTTCCACCACCATCATG 1678
Db 1081 GCCATTTCATATTTCCACCATACACTTCGGCCAGTCTCAGAGCATCTCATTTCCACCACCATCATG 1140
QY 1679 ACT 1681
Db 1141 ACT 1143

RESULT 2
US-09-920-455-250
; Sequence 250, Application US/09920455
; Patent No. US20020168647A1
; GENERAL INFORMATION:
; APPLICANT: Wan4, Tonglong
; APPLICANT: Fan, Liqun

; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY
; FILE REFERENCE: 210121.540
; CURRENT APPLICATION NUMBER: US/09/920,455
; CURRENT FILING DATE: 2001-08-01
; NUMBER OF SEQ ID NOS: 275
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 250
; LENGTH: 337
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc.feature
; LOCATION: 42..43
; OTHER INFORMATION: n = A,T,C or G
; US-09-920-455-250

Query Match 14.1%; Score 319.2; DB 9; Length 337;
Best Local Similarity 98.5%; Pred. No. 3.1e-79;
Matches 321; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

QY 1333 AAACAATTTAAAGATAAGAAAGAAACAATCAGAAGAAACCTGAAATATGATGATGTG 1392
Db 12 AATTTCATTTAAAGATAAGAAAGAAACAATNGAAGAAACCTGAAATATGATGATGTG 71
QY 1393 GAGATTAAAGAGCAGTTGTCCCAAGTATGAATCTCAACTTTCAACAATAGAGGAAAGTA 1452
Db 72 GAGATTAAAGAGCAGTTGTCCCAAGTATGAATCTCAACTTTCAACAATAGAGGAAAGTA 131
QY 1453 GATACAGATGATCGAACTGAAGGCTATTTACGAGCAGACTCACAAAGAGCCCTCCCACTTT 1512
Db 132 GATACAGATGATCGAACTGAAGGCTATTTACGAGCAGACTCACAAAGAGCCCTCCCACTTT 191
QY 1513 GATTCTCAGCAGCCTGCAGTCTTGGAAAGAAAGAGTCAATGATAGTCTATGCTCATCCA 1572
Db 192 GATTCTCAGCAGCCTGCAGTCTTGGAAAGAAAGAGTCAATGATAGTCTATGCTCATCCA 251
QY 1573 CAGGAAGTCTACAAATGAATATGTACCCAGAGGTTGCAAGAAATAATGCCCATTCACATTTCC 1632
Db 252 CAGGAAGTCTACAAATGAATATGTACCCAGAGGTTGCAAGAAATAATGCCCATTCACATTTCC 311
QY 1633 CAGGATACACTCGGCCAGTCCAGACGA 1658
Db 312 CAGGATACACTCGGCCAGTCCAGACGA 337

RESULT 3
US-09-789-561-14
; Sequence 14, Application US/09789561
; Patent No. US20020064818A1
; GENERAL INFORMATION:
; APPLICANT: NI et al.
; TITLE OF INVENTION: 52 Human secreted proteins
; FILE REFERENCE: P2043P1
; CURRENT APPLICATION NUMBER: US/09/789,561
; CURRENT FILING DATE: 2001-02-22
; PRIOR APPLICATION NUMBER: PCT/US00/24008
; PRIOR FILING DATE: 2000-08-31
; PRIOR APPLICATION NUMBER: 60/152,317
; PRIOR FILING DATE: 1999-09-03
; PRIOR APPLICATION NUMBER: 60/152,315
; PRIOR FILING DATE: 1999-09-03
; NUMBER OF SEQ ID NOS: 194
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 14
; LENGTH: 5330
; TYPE: DNA
; ORGANISM: Homo sapiens
; US-09-789-561-14

Query Match 14.1%; Score 319; DB 10; Length 5330;
Best Local Similarity 56.3%; Pred. No. 2.2e-78;
Matches 736; Conservative 0; Mismatches 535; Indels 36; Gaps 6;


```
QY 969 AGCCGTGGGTTGGTGGTTTATATAGCAATTTCCATCATCAGTTTCTGTCCTGCTGGGGT 1028
Db 1372 AGCCGTGGGTTGGTGGTTTATATAGCAATTTCCATCATCAGTTTCTGTCCTGCTGGGGT 1431
QY 1029 TATCTTACTGCTCTCATGAATCGGCTGTTTTCAAATTTCTCTGAGTTTCTTGTGGC 1088
Db 1432 GATCTTGGTCTATCATTAACCAAGGATGCTTCAAAATTCCTCTCATATTCCTTGTGTC 1491
QY 1089 ACTGCGCCCTTGGGACTTTGAGTGTGATGCTTTTACACCTTCTTCCACATCTCTCATGC 1148
Db 1492 ATTAGCTGTATGACAAATGAGTGTGAGCGCTTCTTCTCATCTAGTCTGCCCATCTCAGG 1551
QY 1149 AAGTCACACATATAGTCTATGACCAATGAAGAACCAAGCAATGMAATGAAGAGAGCACT 1208
Db 1552 TGGACATGATCACAGTCCACCAACATGCACA-----TCGGGCAATGACATTC 1596
QY 1209 TTTTCAGTCATCTGCTTCTCTCAAAACATAGAGAAAGTGCCTATTTTGAATTCACGTTGGAA 1268
Db 1597 TCATGGCAATGAATCTAACAAGTTTTCGGAAGAA-----TATCATGCTGTATTTGAA 1647
QY 1269 GGTCTTAACAGCTTAGCAGCCCTGTATTTTCATCTTTCTTGAACATGCTCTCACATT 1328
Db 1648 AGGACTTCTTCTCTAGAGCGATTTACTTGTCTATTTATCATGGAACACTGCGATTAGAA 1707
QY 1429 GATCAAAACATTTAAAGATAAGA---AGAAAAGAAATCAGAGAAACCTGAAATATGTA 1385
Db 1708 CTTTAAGCCTACAAACAAACAGCAAGAAACAGAAATGCTTTATGAAATAGAAACAGAA 1767
QY 1486 TCATCTGAGATTAAGAGCACTTCTCCAAGTATGAATCTCAACTTTTCAACAATCAGGA 1445
Db 1768 AGAATCAACTTATTCGGAAGAAAGCTTTCAGATCAAGTAAACAAATACACAGATTTGTA 1827
QY 1446 CAAAGTAGATACAGATGATCAACTGAAGGCTATTTACGAGCAGACTCAAGAGAGCTTC 1505
Db 1828 CTGCGTTTCACTCAAGCTCTTGCAGAACTGATGACTCGGTGTTTCTGAAAGATCAACT 1887
QY 1506 CCACTTTTATCTCAGGAGCGCTGAGTCTTGGAAAGAAAGAGGTGATGATAGTTCATGC 1565
Db 1888 TAATGAAATTCGAAGTACAGATTTAGAAAGCCCAAGAAATGCCCTCTTAAAAATTAAGT 1947
QY 1566 TCATCTCACAGCAAGTCTACATGATATATGACCAAGAGTGTGCAAGATTAATATCAATTC 1625
Db 1948 TTGTATAGAGAGAGAG-AAAATCAGACCACTTCTCACAGTGAATGGAATATCATACCAT- 2005
QY 1626 ACATTTTCCAGATACACTCGGCCAGTTCAGCAGATCTCATTTCACTCACCATCATGACTAGCA 1685
Db 2006 -CATGAGCATGATTTCCATGCTGCTGCACATACCCAGCCAGGCGAGAACAAAATGTGCT 2064
QY 1686 TCATATTTCTCCATCATCACCAACCAACCAACCAACCAACCAACCAACCAACCAACCAACCA 1745
Db 2065 GAGCAAGCAATTAACCCAGTGGCAGCAAGCAAGCAATTTCTCATTTCCCATGGCCCTCTGCA 2124
QY 1746 CTCTCTGGAGAGGCTGAAGATGCGCGCTGCGGCACTTTGGCCCTGGATGGTGTATATATGG 1805
Db 2125 TTCTGTATCCGCAATCTGAAGAAACAGGAAATAGCTAATATAGCCCTGGATGCTCATATGG 2184
QY 1806 TGATGGCTTGCACAAATTTACAGCATGCGCTAGCAATTTGGTGGCTGCTTTTACTCAAGAGCTT 1865
Db 2185 GGATGGATCCACAACTTCAGTGTATGGCTGCGCAATTTGGTGGAGCTTTTCACTGCTGATTT 2244
QY 1866 ATCAAGTGTGTTTAAATGATCTCTGTTGCTGTGTTCTGTCTCATGAGTTGGCTATGAATTAGG 1925
Db 2245 GACAGGAGGAATCAGTACTTCTATAGCGCTCTTCTGTCTCATGAATGCGCAATTAATAGG 2304
QY 1926 TGACTTTTCTGCTTCTTAAAGGCTGGCATGAGGTTTAAGCAGCGCTGCTCTCTTATATATGC 1985
Db 2405 AGATTTTTCGAGTTCTTCTTAAAGCAGCGCATGACTGTAAAGCAAGCAATTTCTATATCAACT 2464
QY 1986 ATTTCTCAGCCATCTGGCGTATCTTGAATGGCAACAGGAATTTTCACTTCTGCTCATATGC 2045
Db 2465 CCTCTCTGCCATGATGGCTTACATAGGCATAGGCATAGGCACAGCTCTCTCTAGATGCT 2424
```

```
QY 2046 TGAAAATGTTCTTATGTTGGATATTTGCACATTTACTGCTGGCTATTTTCATGATGTTGCTCT 2105
Db 2425 CAATAACATCACACTTTGGATCTTTGCAGTCACTGCTGAGCATGTTCTCTCTATGTAACCTT 2484
QY 2106 GGTGTATATGTTACCTGAAATGCTGCACAAATGATG-----CTAGTGACCATCGATGTAG 2159
Db 2485 GGTGGATATGCTTCCAGAAATGTTGCATGGTGTGACAATGAGAAACATGGCTTTTG 2544
QY 2160 CCGCTGGGGTATTTCTTTTACAGAAATGCTGGATGCTTTGGGTTTGGGAATATGTT 2219
Db 2545 TCTGTGGGCAATTCATCTTCAGAAATTTAGGATTTGCTTTGGATTTGGCAATATGCT 2604
QY 2220 ACTTATTTCCATATTTGAACATAAAATCGTGTGTTTCGTATATAAAATTTCT 2266
Db 2605 GGTGATTTGCCCTCTATGATGAAGATAAAATTCGTTTGACATCCACATTTT 2651

RESULT 4
US-09-880-107-1649
: Sequence 1649, Application US/09880107
: Patent No. US20020142981A1
: GENERAL INFORMATION:
: APPLICANT: Horne, Darci T.
: APPLICANT: Vockley, Joseph G.
: APPLICANT: Scherf, Uwe
: APPLICANT: Gene Logic, Inc.
: TITLE OF INVENTION: Gene Expression Profiles in Liver Cancer
: FILE REFERENCE: 4921-5028-WO
: CURRENT APPLICATION NUMBER: US/09/880,107
: CURRENT FILING DATE: 2001-06-14
: PRIOR APPLICATION NUMBER: US 60/211,379
: PRIOR FILING DATE: 2000-06-14
: PRIOR APPLICATION NUMBER: US 60/237,054
: PRIOR FILING DATE: 2000-10-02
: NUMBER OF SEQ ID NOS: 3950
: SOFTWARE: Patent In Ver. 2.1
: SEQ ID NO 1649
: LENGTH: 4573
: TYPE: DNA
: ORGANISM: Homo sapiens
: FEATURE:
: OTHER INFORMATION: Genbank Accession No. US20020142981A1 D31887
US-09-880-107-1649
```

```
Query Match 5.6%; Score 127.6; DB 10; Length 4573;
Best Local Similarity 59.6%; Pred. No. 4.7e-25;
Matches 236; Conservative 0; Mismatches 154; Indels 6; Gaps 1;

QY 1732 CACAGCCAGCGTACTCTCGGGAGGAGCTGAAAGATGCCGGCTGCCCACTTTGGCCTGG 1791
Db 1074 CAGACTGCTTGTCTACTGGCTGAAAGGTGTCGCCACTCTGTGATATCGGCACTCTGGCCTGG 1133
QY 1792 ATGGTGATTAATGGGTATGCGCTGCACAAATTTACGCGATGGCTAGCAATTTGGTGCTGCT 1851
Db 1134 ATGATCACTCTGAGGACGCGCTCCACAAATTTTCATCGATGGCTGCCCACTGCTGCTTCC 1193
QY 1852 TTTACTGAAGGCTTATCAAGTGGTTTAACTACTTCTGTTGCTGTTCTGTCATCAGTTG 1911
Db 1194 TTCACTGTCTAGTTTTCCAAGGCAATCAGCACTCGTGGGCCATCTCTGTGAGGAGTTC 1253
QY 1912 CCTCATGAATTAGTGACTTTTCTGTACTTAAAGGCTGGCATGACCCCTTAAGCAGCT 1971
Db 1254 CCACATGAGCTAGAGAGACTTTTGTCTATCTGCTCAAGCTGGGATGAGCATCCACAAGCT 1313
QY 1972 GTCTTTTATAATGCAATGTGACGCCATGCTGCGGTATCTTTGGAAATGCCAACAGGAATTTTC 2031
Db 1314 CTCTTCTTCAACTTCTTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1373
QY 2032 ATTTGTCATATGCTCAAAATGTTTCTATGTCGATATTTTGGCACTTTTACTGCTGGCTTATTC 2091
Db 1374 GCGGG-----CAGCCACTTCTCTGCCAACGGATTTTGGCGCTAGCTGGAGGAATGTTTC 1427
QY 2092 ATGTATGTTGCTCTGTTGATATGCTACCTGAAATG 2127
```



```

: ORGANISM: Homo sapiens
: FEATURE:
: OTHER INFORMATION: MAP TO AL079301.14
: OTHER INFORMATION: EXPRESSED IN HBL100, SIGNAL = 7.8
: OTHER INFORMATION: EXPRESSED IN FETAL LIVER, SIGNAL = 5.2
: OTHER INFORMATION: EXPRESSED IN HELA, SIGNAL = 8.2
: OTHER INFORMATION: EXPRESSED IN BRAIN, SIGNAL = 8.4
: OTHER INFORMATION: EXPRESSED IN PLACENTA, SIGNAL = 4.7
: OTHER INFORMATION: EXPRESSED IN HEART, SIGNAL = 4.3
: OTHER INFORMATION: EXPRESSED IN ADULT LIVER, SIGNAL = 5.6
: OTHER INFORMATION: EXPRESSED IN LUNG, SIGNAL = 6.2
: OTHER INFORMATION: EXPRESSED IN BONE MARROW, SIGNAL = 7.8
: OTHER INFORMATION: EXPRESSED IN BT474, SIGNAL = 7
: OTHER INFORMATION: EST_HUMAN_HIT: BE011604.1, EVALU6 1.10e-01
: OTHER INFORMATION: NT_HIT: g19628932, EVALU6 1.10e-01
US-09-864-761-17051

Query Match      2.7%; Score 62.2; DB 10; Length 66%;
Best Local Similarity 59.2%; Pred. No. 2.3e-07;
Matches 106; Conservative 0; Mismatches 73; Indels 0; Gaps 0;

QY 271 CATATACCATCAGCAGCAGCATCACTCAGACGAGCATCACTCACTCACTGAGCT 330
    ||| ||||| ||||| ||| ||| ||| ||||| ||| ||||| ||| ||||| |||
Db 300 CATCACCACCATCAGCAGCAGCATCACTCAGCAGCAGCATCACTCAGCAGCAGC 359
    ||| ||||| ||||| ||| ||| ||| ||||| ||| ||||| ||| ||||| |||
QY 331 CACTCAGACCATGAGCATCACTCAGCAGCAGCATCACTCTGACCATGATCATCACTCT 390
    ||| ||||| ||||| ||| ||| ||| ||||| ||| ||||| ||| ||||| |||
Db 360 CACCATCAGCATGAGCAGCAGCAGCATCACTCAGCAGCAGCATCACTGACCATGATCATCACT 419
QY 391 CACCATCAATCATGCTGCTCTGTTGTAATAATAGCGAAAGCTCTTTGGCCAGACCATGA 449
    ||||| ||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 420 CACCATTATCAGCATCATTTATGACACCATCATCACTCAGCAGCAGCATCACTGAGCAGCAGCATCA 478

RESULT 7
US-09-286-488-64/c
: Sequence 64, Application US/09286488
: Patent No. US20020169136A1
: GENERAL INFORMATION:
: APPLICANT: Reed, Steven G.
: APPLICANT: Lodes, Michael J.
: APPLICANT: Houghton, Raymond L.
: APPLICANT: Sleath, Paul R.
: APPLICANT: McNeill, Patricia D.
: TITLE OF INVENTION: COMPOUNDS AND METHODS FOR THE DIAGNOSIS
: TITLE OF INVENTION: AND TREATMENT OF B. MICROTI INFECTION
: FILE REFERENCE: 210121.426C3
: CURRENT APPLICATION NUMBER: US/09/286,488
: CURRENT FILING DATE: 1999-04-05
: NUMBER OF SEQ ID NOS: 83
: SOFTWARE: FastSeq for Windows Version 3.0
: SEQ ID NO 64
: LENGTH: 285
: TYPE: DNA
: ORGANISM: Babesia microti
US-09-286-488-64

Query Match      2.6%; Score 59.8; DB 9; Length 285;
Best Local Similarity 65.2%; Pred. No. 6.1e-07;
Matches 88; Conservative 0; Mismatches 47; Indels 0; Gaps 0;

QY 268 ATCCATATACACCATGAGCATCACTCAGACGAGCATCACTCTGAGCATGATCATGAG 327
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 192 AACCACTTCCACTAGGCCACCACTTCTACTAGGCCACCACTTCTACTAGGCCACCACTTCTAC 133
QY 328 CGTCACTAGACCATGAGCATCACTCAGACGAGCATCACTCTGAGCATGATCATGAG 387
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 132 CTTCACTAGGCCACCACTTCTACTAGGCCACCACTTCTACTAGGCCACCACTTCTAC 73
QY 388 TCTCACCATAATCAT 402
    ||| ||| |||
Db 72 TAGGCCACCACTGCTT 58

RESULT 9
US-09-286-488-65/c
: Sequence 65, Application US/09286488
: Patent No. US20020169136A1
: GENERAL INFORMATION:
: APPLICANT: Reed, Steven G.
: APPLICANT: Lodes, Michael J.
: APPLICANT: Houghton, Raymond L.
: APPLICANT: Sleath, Paul R.
: APPLICANT: McNeill, Patricia D.
: TITLE OF INVENTION: COMPOUNDS AND METHODS FOR THE DIAGNOSIS
: TITLE OF INVENTION: AND TREATMENT OF B. MICROTI INFECTION
: FILE REFERENCE: 210121.426C3
: CURRENT APPLICATION NUMBER: US/09/286,488
: CURRENT FILING DATE: 1999-04-05
: NUMBER OF SEQ ID NOS: 83
: SOFTWARE: FastSeq for Windows Version 3.0
: SEQ ID NO 65
: LENGTH: 342
: TYPE: DNA
: ORGANISM: Babesia microti
US-09-286-488-65

Query Match      2.6%; Score 59.8; DB 9; Length 342;
Best Local Similarity 65.2%; Pred. No. 6.8e-07;
Matches 88; Conservative 0; Mismatches 47; Indels 0; Gaps 0;

QY 268 ATCCATATACACCATGAGCATCACTCAGACGAGCATCACTCTGAGCATGATCATGAG 327
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 210 AACCACTTCCACTAGGCCACCACTTCTACTAGGCCACCACTTCTACTAGGCCACCACTTCTAC 151
```

```

RESULT 8
US-09-737-178-64/c
: Sequence 64, Application US/09737178
: Patent No. US20010029295A1
: GENERAL INFORMATION:
: APPLICANT: Reed, Steven G.
: APPLICANT: Lodes, Michael J.
: APPLICANT: Houghton, Raymond L.
: APPLICANT: Sleath, Paul R.
: APPLICANT: McNeill, Patricia D.
: APPLICANT: Homer, Mary
: APPLICANT: Secrist, Heather
: TITLE OF INVENTION: COMPOUNDS AND METHODS FOR THE DIAGNOSIS
: TITLE OF INVENTION: AND TREATMENT OF B. MICROTI INFECTION
: FILE REFERENCE: 210121.426C9
: CURRENT APPLICATION NUMBER: US/09/737,178
: CURRENT FILING DATE: 2000-12-13
: NUMBER OF SEQ ID NOS: 144
: SOFTWARE: FastSeq for Windows Version 3.0
: SEQ ID NO 64
: LENGTH: 285
: TYPE: DNA
: ORGANISM: Babesia microti
US-09-737-178-64

Query Match      2.6%; Score 59.8; DB 10; Length 285;
Best Local Similarity 65.2%; Pred. No. 6.1e-07;
Matches 88; Conservative 0; Mismatches 47; Indels 0; Gaps 0;

QY 268 ATCCATATACACCATGAGCATCACTCAGACGAGCATCACTCTGAGCATGATCATGAG 327
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 192 AACCACTTCCACTAGGCCACCACTTCTACTAGGCCACCACTTCTACTAGGCCACCACTTCTAC 133
QY 328 CGTCACTAGACCATGAGCATCACTCAGACGAGCATCACTCTGAGCATGATCATGAG 387
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 132 CTTCACTAGGCCACCACTTCTACTAGGCCACCACTTCTACTAGGCCACCACTTCTAC 73
QY 388 TCTCACCATAATCAT 402
    ||| ||| |||
Db 72 TAGGCCACCACTGCTT 58

RESULT 9
US-09-286-488-65/c
: Sequence 65, Application US/09286488
: Patent No. US20020169136A1
: GENERAL INFORMATION:
: APPLICANT: Reed, Steven G.
: APPLICANT: Lodes, Michael J.
: APPLICANT: Houghton, Raymond L.
: APPLICANT: Sleath, Paul R.
: APPLICANT: McNeill, Patricia D.
: TITLE OF INVENTION: COMPOUNDS AND METHODS FOR THE DIAGNOSIS
: TITLE OF INVENTION: AND TREATMENT OF B. MICROTI INFECTION
: FILE REFERENCE: 210121.426C3
: CURRENT APPLICATION NUMBER: US/09/286,488
: CURRENT FILING DATE: 1999-04-05
: NUMBER OF SEQ ID NOS: 83
: SOFTWARE: FastSeq for Windows Version 3.0
: SEQ ID NO 65
: LENGTH: 342
: TYPE: DNA
: ORGANISM: Babesia microti
US-09-286-488-65

Query Match      2.6%; Score 59.8; DB 9; Length 342;
Best Local Similarity 65.2%; Pred. No. 6.8e-07;
Matches 88; Conservative 0; Mismatches 47; Indels 0; Gaps 0;

QY 268 ATCCATATACACCATGAGCATCACTCAGACGAGCATCACTCTGAGCATGATCATGAG 327
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 210 AACCACTTCCACTAGGCCACCACTTCTACTAGGCCACCACTTCTACTAGGCCACCACTTCTAC 151
```


Query Match

```

; NUMBER OF SEQ ID NOS: 83
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 58
; LENGTH: 358
; TYPE: DNA
; ORGANISM: Babesia microti
US-09-286-488-58

```

Query Match	2.68;	Score 59.8;	DB 9;	Length 358;
Best Local Similarity	65.24;	Pred. No. 7e-07;		
Matches 88;	Conservative	0;	Mismatches 47;	Indels 0;
				Gaps 0;

APPLICANT: SECTIS, REALTE
TITLE OF INVENTION: COMPOUNDS AND METHODS FOR THE DIAGNOSIS
TITLE OF INVENTION: AND TREATMENT OF B. MICROTI INFECTION

Search completed: November 28, 2002, 09:24:26
Job time : 157 secs

RESULT 15
US-09-286-488-58/c
: Sequence 58, Application US/09286488

GenCore version 5.1.3
Copyright (c) 1993 - 2002 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: November 27, 2002, 22:23:30 ; Search time 439 Seconds
(without alignments)
11634.470 Million cell updates/sec

Title: US-09-642-034-4

Perfect score: 2268
Sequence: 1 atggcgaggagattatctgt.....tggttcgtataaattttctag 2268

Scoring table: IDENTITY_NUC

Gapop 10.0 , Gapext 1.0

Searched: 2185239 seqs, 1125999159 residues

Total number of hits satisfying chosen parameters: 4370478

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : N_Geneseq_101002.*
1: /SID52/gcgdata/geneseq/geneseq-emb1/NA1980.DAT.*
2: /SID52/gcgdata/geneseq/geneseq-emb1/NA1981.DAT.*
3: /SID52/gcgdata/geneseq/geneseq-emb1/NA1982.DAT.*
4: /SID52/gcgdata/geneseq/geneseq-emb1/NA1983.DAT.*
5: /SID52/gcgdata/geneseq/geneseq-emb1/NA1984.DAT.*
6: /SID52/gcgdata/geneseq/geneseq-emb1/NA1985.DAT.*
7: /SID52/gcgdata/geneseq/geneseq-emb1/NA1986.DAT.*
8: /SID52/gcgdata/geneseq/geneseq-emb1/NA1987.DAT.*
9: /SID52/gcgdata/geneseq/geneseq-emb1/NA1988.DAT.*
10: /SID52/gcgdata/geneseq/geneseq-emb1/NA1989.DAT.*
11: /SID52/gcgdata/geneseq/geneseq-emb1/NA1990.DAT.*
12: /SID52/gcgdata/geneseq/geneseq-emb1/NA1991.DAT.*
13: /SID52/gcgdata/geneseq/geneseq-emb1/NA1992.DAT.*
14: /SID52/gcgdata/geneseq/geneseq-emb1/NA1993.DAT.*
15: /SID52/gcgdata/geneseq/geneseq-emb1/NA1994.DAT.*
16: /SID52/gcgdata/geneseq/geneseq-emb1/NA1995.DAT.*
17: /SID52/gcgdata/geneseq/geneseq-emb1/NA1996.DAT.*
18: /SID52/gcgdata/geneseq/geneseq-emb1/NA1997.DAT.*
19: /SID52/gcgdata/geneseq/geneseq-emb1/NA1998.DAT.*
20: /SID52/gcgdata/geneseq/geneseq-emb1/NA1999.DAT.*
21: /SID52/gcgdata/geneseq/geneseq-emb1/NA2000.DAT.*
22: /SID52/gcgdata/geneseq/geneseq-emb1/NA2001A.DAT.*
23: /SID52/gcgdata/geneseq/geneseq-emb1/NA2001B.DAT.*
24: /SID52/gcgdata/geneseq/geneseq-emb1/NA2002.DAT.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	2268	100.0	3461	24	ABK92205
2	2268	100.0	3461	24	ABK92205
3	2264.8	99.9	2776	22	AA13480
4	2220.4	97.9	3523	23	ABV25747
5	2198	96.9	3461	22	AA13479
6	1263	55.7	1310	19	AAT99070
7	1263	55.7	2404	19	AAT99071
8	1259.8	55.5	1310	17	AAT33219
9	1259.8	55.5	2404	17	AAT33220

10	1141.4	50.3	1193	21	AAF16156	Human prostate can
11	787.4	34.7	20778	22	AAK79819	Human immune/haema
12	619.6	27.3	899	24	ABA92300	Mouse breast cancer
13	549.4	24.2	551	22	AA13487	Human LIV-1-164647
14	509.4	22.5	1597	20	AA233622	Human breast tumou
15	390.4	17.2	424	22	AA19686	Human breast cancer
16	380.4	16.8	419	22	AA122452	Human breast cancer
17	319.2	14.1	337	24	ABK54054	Human head and nec
18	319	14.1	3309	23	ABV23278	Human prostate exp
19	319	14.1	3309	23	ABV29124	Human prostate exp
20	319	14.1	5322	24	ABN59760	Novel human coding
21	319	14.1	5330	22	AAF91861	Human secreted pro
22	262.4	11.6	423	22	AA123559	Human breast cancer
23	251.4	11.1	469	22	AA101016	Human breast cancer
24	247	10.9	736	22	AAK92049	Human cDNA 5'-end
25	247	10.9	736	22	AAK93710	Human cDNA clone r
26	247	10.9	2863	22	AAK94381	Human full-length
27	240.8	10.6	413	22	AA19430	Human breast cancer
28	240.8	10.6	455	22	AA109043	Human breast cancer
29	224.8	9.9	406	23	ABV60690	Human prostate exp
30	206.6	9.1	643	22	AA18330	Human breast cancer
31	189.4	8.4	405	22	AA18310	Human breast cancer
32	157.4	6.9	233	23	ABV38837	Human prostate exp
33	157	6.9	176	22	AA117684	Human breast cancer
34	147.2	6.5	567	23	ABV08950	Human prostate exp
35	146.4	6.5	2055	23	ABL11749	Drosophila melanog
36	146	6.4	193	22	AA110609	Human breast cancer
37	144	6.3	4370	23	ABL11748	Drosophila melanog
38	143.8	6.3	183	22	AA16763	Human breast cancer
39	141.4	6.2	409	22	AA108417	Human breast cancer
40	140.4	6.2	207	22	AA18232	Human breast cancer
41	140.4	6.2	378	22	AA126162	Human breast cancer
42	128.4	5.7	554	22	AA113583	Human breast cancer
43	127.6	5.6	4573	20	AAK33946	Human HCMV Inducib
44	127.6	5.6	4573	24	ABN95151	Gene #1649 used to
45	127.6	5.6	4631	24	ABK63926	cDNA encoding huma

ALIGNMENTS

RESULT 1

ABK92205	ABK92205 standard; DNA; 3461 BP.
ID	ABK92205
XX	AC
XX	ABK92205;
XX	15-AUG-2002 (first entry)
DT	Prostate cancer-associated DNA sequence #91.
DE	Prostate cancer; prostate tumour tissue; human; mammal; cytostatic;
XX	gene therapy; gene; ds.
KW	Mammalia.
KW	WO200230268-A2.
XX	18-APR-2002.
XX	12-OCT-2001; 2001WO-US32045.
PF	13-OCT-2000; 2000US-0687576.
XX	08-DEC-2000; 2000US-0733288.
PR	08-DEC-2000; 2000US-0733288.
PR	08-DEC-2000; 2000US-0733742.
PR	24-JAN-2001; 2001US-263957P.
PR	16-MAR-2001; 2001US-276791P.
PR	16-MAR-2001; 2001US-276888P.
PR	06-APR-2001; 2001US-281922P.
PR	24-APR-2001; 2001US-286214P.
PR	30-APR-2001; 2001US-0847046.
PR	04-MAY-2001; 2001US-288589P.
XX	

PA	(EUSB-) EOS BIOTECHNOLOGY INC..
XX	Gish KC, Mack DH, Wilson KE, Afar D, Hevezi P;
XX	WPT: 2002-471335/50.
DK	P-PSDB; ABG61889.
XX	Detecting a prostate cancer-associated transcript in a cell in a
PT	patient, useful for diagnosing prostate cancer (PC) or screening
PT	modulators of PC, by determining if prostate cancer-associated genes
PT	are expressed in a prostate tissue
XX	
PS	Claim 22: Page 374-375; 436pp; English.
XX	The present invention relates to methods of detecting a prostate
CC	cancer-associated transcript in a cell from a patient. The method
CC	comprises contacting a biological sample from the patient with
CC	prostate cancer-associated polynucleotides (designated PC genes) that
CC	selectively hybridize to a sequence that is at least 80% identical
CC	to them. The prostate cancer-associated polynucleotide sequences
CC	are differentially expressed in prostate tumour tissue or in
CC	prostate cancer and are derived from the tissues of various
CC	organisms such as humans or other mammals (e.g. mice, sheep and dogs).
CC	The methods of the invention are useful for diagnosing and treating
CC	prostate cancer in mammals. The prostate cancer-associated genes are
CC	useful for diagnosing or treating prostate cancer, as well as for
CC	identifying modulators of prostate cancer or agents that inhibit
CC	prostate cancer. The nucleic acid sequences are particularly useful
CC	in gene therapy, as a vaccine or in antisense applications.
CC	AHK92115-AHK92263 represent prostate cancer-associated polynucleotide
CC	sequences.
XX	
SQ	Sequence 1461 BP; 1042 A; 686 C; 728 G; 1005 T; 0 other;
Query Match 100.0%; Score 2268; DB 24; Length 1461;	
Best Local Similarity 100.0%; Pred. No. 0;	
Matches 2268; Conservative 0; Mismatches 0; Indels 0; Gaps 0;	
QY	1 ATGCGAGGAAGTATCTGTAACTTTGATCCTGCACCTTTGCCCTCTCTGTCACAATAATCCC
DB	138 ATGCGAGGAAGTATCTGTAACTTTGATCCTGCACCTTTGCCCTCTCTGTCACAATAATCCC
QY	61 CTTCATGUAATTAAGAAGCAGTGCTTTCCCCAGACCACCTGAGAAAAATTAGTCTGAAATTCG
DB	198 CTTCATGUAATTAAGAAGCAGTGCTTTCCCCAGACCACCTGAGAAAAATTAGTCTGAAATTCG
QY	121 GAATCTGGCAATTAATGTTGACTTTGGCAATTTCCACAGCGCAATATCATCTACACAGCTTT
DB	258 GAATCTGGCAATTAATGTTGACTTTGGCAATTTCCACAGCGCAATATCATCTACACAGCTTT
QY	181 TTCTPACCGCTATCGAGAAAATAATCTTTCTGACTTTGAAGGTTTCAGAAAATTAATCTCAA
DB	318 TTCTPACCGCTATCGAGAAAATAATCTTTCTGACTTTGAAGGTTTCAGAAAATTAATCTCAA
QY	241 AATATAGGCATAGATAAGATTAAAGAAATCCATATACCATGACCACGACCATCATCTCA
DB	378 AATATAGGCATAGATAAGATTAAAGAAATCCATATACCATGACCACGACCATCATCTCA
QY	301 CACACAGCAGCATCACTTCAGACCATGAGCGTCACTTCAGACCATGAGCATCACTCACACAC
DB	438 CACACAGCAGCATCACTTCAGACCATGAGCGTCACTTCAGACCATGAGCATCACTCACACAC
QY	361 GAGCATCACTCTCACCATGATCATCACTCTCACCATTAATCATGCTTCCTTCAANAAT
DB	498 GAGCATCACTCTCACCATGATCATCACTCTCACCATTAATCATGCTTCCTTCAANAAT
QY	421 AGCGAAAAAGCTCTTTTGGCCAGACCATGACTTCAGATAGTTCAGGTAAAGATTCCTTAACAC
DB	558 AGCGAAAAAGCTCTTTTGGCCAGACCATGACTTCAGATAGTTCAGGTAAAGATTCCTTAACAC
QY	481 AGCCAGGGGAAGAGCTCACCAGCAGAACATGCCAGTGGTAGAAGGAATGTCACAGAC
DB	618 AGCCAGGGGAAGAGCTCACCAGCAGAACATGCCAGTGGTAGAAGGAATGTCACAGAC
QY	541 AGTGTTAGTGTCTAGTGAAGTGACCTCAACTGTGTACAACACTGTCTCTGAAGAACTCAC
DB	678 AGTGTTAGTGTCTAGTGAAGTGACCTCAACTGTGTACAACACTGTCTCTGAAGAACTCAC
QY	601 TTTCTAGACACAATAGAGACTCCAAGACCTGGAAAACCTTTCCCCCAAGACATGAAGCAGC
DB	738 TTTCTAGACACAATAGAGACTCCAAGACCTGGAAAACCTTTCCCCCAAGACATGAAGCAGC
QY	661 TTCACTCCACCAGTGTCTACATCAAGAGCCGGGTGAGCCGGCTGGCTGGTAGGAAAAACA
DB	798 TTCACTCCACCAGTGTCTACATCAAGAGCCGGGTGAGCCGGCTGGCTGGTAGGAAAAACA
QY	721 AATGAATCTGTGAGTGAGCCCGGAAAAGGCTTTATGTATTCCAGAAACACAATCAAAAT
DB	858 AATGAATCTGTGAGTGAGCCCGGAAAAGGCTTTATGTATTCCAGAAACACAATCAAAAT
QY	781 CCTCAGGAGTGTTTCAATCATCAAGCTTACTGACATCTCATGGCATGGGATCCAGGTT
DB	918 CCTCAGGAGTGTTTCAATCATCAAGCTTACTGACATCTCATGGCATGGGATCCAGGTT
QY	841 CGCTGAATGCAACAGAGTTCAACTATCTCTGTCAGGCATCATCAACCAATTTGATGCT
DB	978 CGCTGAATGCAACAGAGTTCAACTATCTCTGTCAGGCATCATCAACCAATTTGATGCT
QY	901 ACATCTCTCTCATTCATACAGTCAAAAAGAGGCTGAATCCCTCCAAAGACCTATTTC
DB	1038 AGATCTTGTCTGATTCATACAGTCAAAAAGAGGCTGAATCCCTCCAAAGACCTATTTC
QY	961 TTACAAATAGCCCTGGTGGTGTATATAGCCATTTCCATCATCAGTTCTCTGCTCTG
DB	1098 TTACAAATAGCCCTGGTGGTGTATATAGCCATTTCCATCATCAGTTCTCTGCTCTG
QY	1021 CTGGGGTTATCTTAGTCCCTCTCATGAATCGGGTGTTTTTCAAAATTTCTCTCGATTTTC
DB	1158 CTGGGGTTATCTTAGTCCCTCTCATGAATCGGGTGTTTTTCAAAATTTCTCTCGATTTTC
QY	1081 CTGTGTGGCACTGGCCGTTGGGACCTTTGAGTGGTGATGCTTTTTTACACCTCTTCCACAT
DB	1218 CTGTGTGGCACTGGCCGTTGGGACCTTTGAGTGGTGATGCTTTTTTACACCTCTTCCACAT
QY	1141 TCCTATGCAAGTCACACCATAGTCATAGCCATGAAGAACACAGCAATGGAATGAAAAGA
DB	1278 TCCTATGCAAGTCACACCATAGTCATAGCCATGAAGAACACAGCAATGGAATGAAAAGA
QY	1201 GGACCACTTTTCAGTCATCTGTCTTCTCAAAACATAGAGAAAGTGCCATTTTGTATTTC
DB	1338 GGACCACTTTTCAGTCATCTGTCTTCTCAAAACATAGAGAAAGTGCCATTTTGTATTTC
QY	1261 AGTGGAAGGGTCTAACAGCTCTAGGAGCGCTGTATTTCATGTTCTTGTGTAACATGTC
DB	1398 AGTGGAAGGGTCTAACAGCTCTAGGAGCGCTGTATTTCATGTTCTTGTGTAACATGTC
QY	1321 CTCACATTTGATCAAAACAATTTAAAGATGAAGAAAAAGAAATCAGAGAAACCTGAAAT
DB	1458 CTCACATTTGATCAAAACAATTTAAAGATGAAGAAAAAGAAATCAGAGAAACCTGAAAT
QY	1381 CATGATGATGTGGAGTTAAGAAGCAGTTGTCCAAGTATGAATCTCAACTTTTCAACAAAT
DB	1518 CATGATGATGTGGAGTTAAGAAGCAGTTGTCCAAGTATGAATCTCAACTTTTCAACAAAT
QY	1441 GAGGAAAG

QY	1621	CATTTCACATTTTCACGATACACTCGGCCAGTCTCAGACGATCTCATTCACACACATCATGAC	1680
Db	1758	CAATTCACATTTCCACGATACACTCGGCCAGTCTCAGACGATCTCATTCACACACATCATGAC	1817
QY	1681	TACCATCATATTTCTCCATCATCACCACCAACCAACCACTCCCTCACTACATTCACACAGTCAG	1740
Db	1818	TACCATCATATTTCTCCATCATCACCACCAACCAACCACTCCCTCACTACATTCACACAGTCAG	1877
QY	1741	CGCTACTCTCGGAGGAGCTGAAGATGCGGCGTCTGGCCACTTTGGCTCGGATCGTGATA	1800
Db	1878	CGCTACTCTCGGAGGAGCTGAAGATGCGGCGTCTGGCCACTTTGGCTCGGATCGTGATA	1937
QY	1801	ATGGGTGATGGCTGTCACAAATTTTCACGATGGCTTAGCAATTTGGTGGCTGCTTTTACTCAA	1860
Db	1938	ATGGGTGATGGCTGTCACAAATTTTCACGATGGCTTAGCAATTTGGTGGCTGCTTTTACTCAA	1997
QY	1861	GGCTATTCAAGTGGTTTAACTACTTCTGTGTCTGTCTGTCTGTCTGTCTGTCTGTCTTTAT	1920
Db	1998	GGCTATTCAAGTGGTTTAACTACTTCTGTGTCTGTCTGTCTGTCTGTCTGTCTGTCTTTAT	2057
QY	1921	TTAGGTGACTTTTCT	1980
Db	2058	TTAGGTGACTTTTCT	2117
QY	1981	ANTGCATTGTTCAGCCATGCTGGCGTATCTTGGAAATGGCAACAGGAATTTTCATTCGTCAT	2040
Db	2118	AATGCATTGTTCAGCCATGCTGGCGTATCTTGGAAATGGCAACAGGAATTTTCATTCGTCAT	2177
QY	2041	TATGCTGMAAAATGTTTCTATGTGCGATATTTGCACTTACTGCTGGCTTATATGCAATGTTT	2100
Db	2178	TATGCTGMAAAATGTTTCTATGTGCGATATTTGCACTTACTGCTGGCTTATATGCAATGTTT	2237
QY	2101	GTCTGCTTCTATATGCTACCTCAAAATGCTGCAACATGCTAGTGACCATGGATGTAGC	2160
Db	2238	GTCTGCTTCTATATGCTACCTCAAAATGCTGCAACATGCTAGTGACCATGGATGTAGC	2297
QY	2161	CGCTGGGGGATTTTCTTTTACAGAAATGCTGGCATGCTTTTGGGTTTGGAAATTTATGCTTA	2220
Db	2298	CGCTGGGGGATTTTCTTTTACAGAAATGCTGGCATGCTTTTGGGTTTGGAAATTTATGCTTA	2357
QY	2221	CTTATTTTCCATATTTCACATATAATTCGTTTTCGTAATAAATTTCTAG	2268
Db	2358	CTTATTTTCCATATTTCACATATAATTCGTTTTCGTAATAAATTTCTAG	2405
RESULT 2			
AHA92299			
LD	AHA92299 standard; cDNA; 3461 BP.		
XX	AHA92299;		
AC			
DT	10-JUN-2002 (first entry)		
DT			
XX	Human breast cancer 4 gene (BCR4) cDNA.		
DE			
XX			
KW	BCR4; human; breast cancer 4 gene; prostate cancer; diagnosis;		
KW	gene therapy; vaccine; biochip; chromosome 18q12; gene; ss.		
XX			
OS	Homo sapiens.		
XX			
FI	Key Location/Qualifiers		
FT	CDS 138...2405		
FT	/*tag= a		
FT	/product= "BCR4"		
FT	sig_peptide 138...197		
FT	/*tag= b		
FT	mat_peptide 198...2402		
FT	/*tag= C		
XX			
PN	W0200216939-A2.		
XX			
PD	28-FEB-2002.		
XX			

PF	20-AUG-2001; 2001WO-US25997.
XX	
PR	18-AUG-2000; 2000US-0642034.
PR	08-DEC-2000; 2000US-0733320.
XX	
PA	(EOSB-) EOS BIOTECHNOLOGY INC.
XX	
PI	Mack D, Gish KC, Wilson KE;
XX	
DR	WPI; 2002-242033/29.
DR	P-PSDB; AAM51198.
XX	
PT	Screening for drugs that affect expression of the breast cancer 4 gene
PT	or its fragments, use of these to treat prostate and breast cancer, and
PT	diagnosing these diseases -
XX	
PS	Claim 27; Fig 1A-B; 83pp; English.
XX	
CC	The presence sequence is that of cDNA identified for the human
CC	breast cancer 4 gene (BCR4) on chromosome 18 (cytoband 18q12).
CC	The gene was identified as being up-regulated in specific cancer
CC	types in an oligonucleotide microarray interrogated with cRNAs
CC	derived from multiple tissues. The gene encodes the protein given
CC	in AAM51198. The invention provides claimed methods of using BCR4
CC	polynucleotides, polypeptides and antibodies to: screen drug
CC	candidates; screen bioactive agents capable of binding to BCR4;
CC	evaluate the effect of a prostate cancer and/or breast cancer
CC	drug; diagnose prostate cancer or breast cancer; screen for a
CC	bioactive agent capable of interfering with the binding of BCR4 to
CC	an antibody which binds BCR4; neutralise the effect of BCR4; treat
CC	prostate cancer or breast cancer using an inhibitor of BCR4 (e.g.
CC	an antibody); localise a therapeutic moiety to, or treat, prostate
CC	cancer or breast cancer tissue by conjugating an antibody for BCR4
CC	to a cytotoxic agent or radiolabel; inhibit prostate cancer or
CC	breast cancer by administering an antisense molecule to the present
CC	nucleic acid sequence; elicit an immune response by administering
CC	BCR4 or a nucleic acid encoding BCR4; and determine the prognosis
CC	of an individual with prostate cancer or breast cancer by
CC	determining the level of BCR4 in a sample, a high level indicating
CC	a poor prognosis. A biochip comprising BCR4 nucleic acids is also
CC	claimed.
XX	
SQ	Sequence 3461 BP; 1042 A; 686 C; 728 G; 1005 T; 0 other;
	Query Match 100.0%; Score 2268; DB 24; Length 3461;
	Best Local Similarity 100.0%; Pred. No. 0;
	Matches 2268; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Qy	1 ATGGCGAGGAAGTTATCTGTAACTTGATCCTGACCTTTCGCCCTTCTGTGCACAAATCCC 60
Db	138 ATGGCGAGGAAGTTATCTGTAACTTGATCCTGACCTTTCGCCCTTCTGTGCACAAATCCC 197
Qy	61 CTTCATGAACATAAAGCAGCTGCTTTCCCCCAGACCACCTGAGAAAATTAGTCCGAATTGG 120
Db	198 CTTCATGAACATAAAGCAGCTGCTTTCCCCCAGACCACCTGAGAAAATTAGTCCGAATTGG 257
Qy	121 GAATCTGGCAATTAATGTTGACTTGGCAATTTCCACAGGCAATCATCTACAACAGCTT 180
Db	258 GAATCTGGCAATTAATGTTGACTTGGCAATTTCCACAGGCAATCATCTACAACAGCTT 317
Qy	181 TTCTACCGCTATGGAGAAAATAATTTCTTTGTGACGTTGAAGGTTTCAGAAAAATTACTTCAA 240
Db	318 TTCTACCGCTATGGAGAAAATAATTTCTTTGTGACGTTGAAGGTTTCAGAAAAATTACTTCAA 377
Qy	241 AATATAGGCATAGATAGATTAAAGAATCCATATACACCATGACGAGGCATCACTCA 300
Db	378 AATATAGGCATAGATAGATTAAAGAATCCATATACACCATGACGAGGCATCACTCA 437
Qy	301 GACCAGGAGCATCACTCAGACCATTAGCGGTCACCTCAGACCATGAGCATCACTCAGACCA 360
Db	438 GACCAGGAGCATCACTCAGACCATTAGCGGTCACCTCAGACCATGAGCATCACTCAGACCA 497
Qy	361 GAGCATCACTCTGACCATGATCATCACTCTCACCATTAATCATGCTGCTTCGTGTTAAAAAT 420

|||||
498 GAGCATCACTCTGACCATGATCAATCACTCTCACCATAATCATCTGCTTTCTGTTAAAAAT 557
OY AAGCGAAAAGCTCTTTCCGAGACCATCACTCAGATAGTTTCAGTAAAGATCTTACAAAC 480
Db AAGCGAAAAGCTCTTTCCGAGACCATCACTCAGATAGTTTCAGTAAAGATCTTACAAAC 617
OY AGCCAGGGGAAGGAGCTCACCCAGCACAGAACATGCCAGTGGTAGAAGGAATCTCAAGAC 540
Db AGCCAGGGGAAGGAGCTCACCCAGCACAGAACATGCCAGTGGTAGAAGGAATCTCAAGAC 677
OY ACTCTTACTCTGACTGAAGTACCTCAACTGTGTACAACTCTCTCTCTGTAAGGAATCTCAC 600
Db ACTCTTACTCTGACTGAAGTACCTCAACTGTGTACAACTCTCTCTCTGTAAGGAATCTCAC 737
OY TTTTCTAGAGACATACAGACTTCCAGACCTGGAAAACCTCTTCCCAAGATCTTAAGTAC 660
Db TTTTCTAGAGACATACAGACTTCCAGACCTGGAAAACCTCTTCCCAAGATCTTAAGTAC 717
OY TCCACTCCATCCAGTGTACATCAAAAGAGCCGGTGGCTGGCTGGCTGGTGGTGGTGGTGGT 720
Db TCCACTCCATCCAGTGTACATCAAAAGAGCCGGTGGCTGGCTGGCTGGTGGTGGTGGTGGT 797
OY AATCAATCTCTGAGTGGAGCCCAAGAGGCTTTATGTATTCAGAAACACAAATGAAAT 780
Db AATCAATCTCTGAGTGGAGCCCAAGAGGCTTTATGTATTCAGAAACACAAATGAAAT 917
OY CTTCAGGAGTGTCTCAATGCATCAAAAGCTACTCAGCATCTCATGGCATGGCATGGCATGGCAT 840
Db CTTCAGGAGTGTCTCAATGCATCAAAAGCTACTCAGCATCTCATGGCATGGCATGGCATGGCAT 977
OY CCCTGTAATCCAGAGTTCAACTATCTCTGTCAGCCCATCATCAACCAATTCATGCT 1037
Db CCCTGTAATCCAGAGTTCAACTATCTCTGTCAGCCCATCATCAACCAATTCATGCT 900
OY AGATCTTGTCTGATTCATACAACTGAAAGAGGCTGCAATCCCTCCAAAGACTTATTC 960
Db AGATCTTGTCTGATTCATACAACTGAAAGAGGCTGCAATCCCTCCAAAGACTTATTC 1097
OY TTAACAATAGCCCTGGCTTGTGCTTTATAGCCATTTCCATCATCAGTTTCTCTCTCTG 1020
Db TTAACAATAGCCCTGGCTTGTGCTTTATAGCCATTTCCATCATCAGTTTCTCTCTCTG 1157
OY CTGGGGTGTATCTTACTGCTCTCATGAATCGGGTGTCTTCAAAATTTCTCTGAGTTTC 1080
Db CTGGGGTGTATCTTACTGCTCTCATGAATCGGGTGTCTTCAAAATTTCTCTGAGTTTC 1217
OY CTCTGCTGCTGCTGGGCTTTGAGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGT 1140
Db CTCTGCTGCTGCTGGGCTTTGAGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGT 1277
OY TCTCATGCAAGTCAACCATAGTCTATAGCCATGAAGAACAGCAATGCAATGCAATGCAATG 1200
Db TCTCATGCAAGTCAACCATAGTCTATAGCCATGAAGAACAGCAATGCAATGCAATGCAATG 1337
OY GGACCACTTTTTCAGTCACTGCTCTCTCAAAACATAGAAAGTGGCTTATTTTCATTTCC 1260
Db GGACCACTTTTTCAGTCACTGCTCTCTCAAAACATAGAAAGTGGCTTATTTTCATTTCC 1397
OY ACTCTGCAAGCTCTTAACAGCTCTAGAGGCTGTATTTTCATGTTTCTTCTGCAACATCTC 1320
Db ACTCTGCAAGCTCTTAACAGCTCTAGAGGCTGTATTTTCATGTTTCTTCTGCAACATCTC 1457
OY CTTCATGATCAAACTTTTAAAGATGAAGAAAGAAAGAAATCAGAAAGAAATTCAAAT 1380
Db CTTCATGATCAAACTTTTAAAGATGAAGAAAGAAAGAAATCAGAAAGAAATTCAAAT 1517
OY GATGATGATGAGATTAAGAGCAGTTGTCCAAAGTATGATCTCAACTTTTCACAAAT 1440
Db GATGATGATGAGATTAAGAGCAGTTGTCCAAAGTATGATCTCAACTTTTCACAAAT 1577
OY GAGGAGAAAGTATACAGATCATCAACTGAAAGCTTATTTACGAGCAGATCTTACAAAG 1500
|||||

Db 1578 GAGGAGAAAGTATACAGATGATCGAACTGAAGGCTATTTTACGAGCAGACTCACAAGAG 1637
OY CCTCCCACTTTGATTTCTCAGCAGGCTGCAGTCTTTGGAAGAAAGAGTCTCATGATAGCT 1560
Db CCTCCCACTTTGATTTCTCAGCAGGCTGCAGTCTTTGGAAGAAAGAGTCTCATGATAGCT 1697
OY CATGCTCATCCACAGGAAGTCTTACAATGAATATGTATACCCAGAGGTTGCAAGAATAAATGC 1620
Db CATGCTCATCCACAGGAAGTCTTACAATGAATATGTATACCCAGAGGTTGCAAGAATAAATGC 1757
OY CATTCACATTTCCAGCATACACTCGCCAGTCAGACGATCTCATTTCCACCACCATCATCAC 1680
Db CATTCACATTTCCAGCATACACTCGCCAGTCAGACGATCTCATTTCCACCACCATCATCAC 1817
OY TACCATCATATTTCTCCATCATCACACCAACCAACCAACCAACCAACCAACCAACCAACCA 1740
Db TACCATCATATTTCTCCATCATCACACCAACCAACCAACCAACCAACCAACCAACCAACCA 1877
OY CGCTACTCTCGGAGGAGCTGAAAGATCGCCGCTGCGCCACTTTTGGCCCTGGATGGTGATA 1800
Db CGCTACTCTCGGAGGAGCTGAAAGATCGCCGCTGCGCCACTTTTGGCCCTGGATGGTGATA 1937
OY ATGGGTGATGGCTGCACAAATTTACAGCATGGCTTAGCAATTTGGTGTCTTTTACTCAA 1860
Db ATGGGTGATGGCTGCACAAATTTACAGCATGGCTTAGCAATTTGGTGTCTTTTACTCAA 1997
OY GCTTTATCAAGTGGTAACTTCTGTTCTGTTCTGTTCTGTTCTGTTCTGTTCTGTTCTGTTCT 2057
OY TTAGGTGACTTTGCTTCTTACTAAAGCTGCGCATGACCGTTAAGCAGGCTGCTCTTTAT 1980
Db TTAGGTGACTTTGCTTCTTACTAAAGCTGCGCATGACCGTTAAGCAGGCTGCTCTTTAT 2117
OY AATGCATTTGTCAGCCATGCTGGCTATCTTGAATGGCAACAGGAATTTTTCATTTGGTCA 2040
Db AATGCATTTGTCAGCCATGCTGGCTATCTTGAATGGCAACAGGAATTTTTCATTTGGTCA 2177
OY TATGCTGAAAATGTTTCTATGCTGATATTTCCACTTACCTGCTGCTGCTGCTGCTGCTGCT 2100
Db TATGCTGAAAATGTTTCTATGCTGATATTTCCACTTACCTGCTGCTGCTGCTGCTGCTGCT 2237
OY GCTCTGTTGATATGTTACCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 2160
Db GCTCTGTTGATATGTTACCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 2297
OY CGCTGGGGTATTTCTTTTACAGAACTGCTGGGATGCTTTTGGCTTTTGGAAATTTATGTTA 2220
Db CGCTGGGGTATTTCTTTTACAGAACTGCTGGGATGCTTTTGGCTTTTGGAAATTTATGTTA 2357
OY CTTATTTCCATATTTTGAACATAAAATCTGTTTCTGCTATAAATTTCTAG 2268
Db CTTATTTCCATATTTTGAACATAAAATCTGTTTCTGCTATAAATTTCTAG 2405

RESULT 3
AAD13480
ID AAD13480 standard; cdna; 2776 BP.
XX AAD13480;
XX XX
DT 06-NOV-2001 (first entry)
XX Human LIV-1-164647 cdna.
XX
KW Human; LIV-1-164647; cytosolic; estrogen-inducible gene; tumour;
KW cancer; breast; lung; prostate; colon; ovary; kidney; gastric;
KW salivary gland; carcinoma; drug screening; therapy; ss.
XX
OS Homo sapiens.
XX Key
FH Location/Qualities
FT 73..2340


```
QY 1561 CATGCTCAATCCACAGAAAGTCTACAAATGAATATGTACCCAGAGGGTCAAGAAATAAATGC 1620
Db 1562 CATGCTCAATCCACAGAAAGTCTACAAATGAATATGTACCCAGAGGGTCAAGAAATAAATGC 1692
QY 1621 CATTCACATATTCACAGATACACTGCGCCACTCAGAGATCTCATTCACCAACATCATGAC 1680
Db 1693 CATTCACATATTCACAGATACACTGCGCCACTCAGAGATCTCATTCACCAACATCATGAC 1752
QY 1681 TATCAATCAATATTCCTCATATCAATCAACACCAACCAACCAACCAATCTCAGACTCAGACGAC 1740
Db 1753 TATCAATCAATATTCCTCATATCAACACCAACCAACCAACCAATCTCAGACTCAGACGAC 1812
QY 1741 CGTTACTCTCGGAGAGAGCTTGAAGATGCGCGCTGCCACTTTGGCCTGGATGGTATGA 1800
Db 1813 CGCTTACTCTCGGAGAGAGCTTGAAGATGCGCGCTGCCACTTTGGCCTGGATGGTATGA 1872
QY 1801 ATGGGTGATGCGCTGCACAAATTTTACAGGATGGCGCTAGCAATTTGGTCTCTTTACTGAA 1860
Db 1873 ATGGGTGATGCGCTGCACAAATTTTACAGGATGGCGCTAGCAATTTGGTCTCTTTACTGAA 1932
QY 1861 GGCCTTATCAAGTGGTTTAACTACTTCTGTGCTGTCTGTCTGTCATCAGTTTGGCTCATGAA 1920
Db 1933 GGCCTTATCAAGTGGTTTAACTACTTCTGTGCTGTCTGTCTGTCATCAGTTTGGCTCATGAA 1992
QY 1921 TTAGTGTGACTTTGCTGTCTTACTTAAAGGCTGGCATGACCGTTAAGCAGGCTCTGCTTTAT 1980
Db 1993 TTAGTGTGACTTTGCTGTCTTACTTAAAGGCTGGCATGACCGTTAAGCAGGCTCTGCTTTAT 2052
QY 1981 AATGCAATTCACAGCATGCTGGCGTATCTTGAATGGCAACAGCAATTTTCATTTGCTCAT 2040
Db 2053 AATGCAATTCACAGCATGCTGGCGTATCTTGAATGGCAACAGCAATTTTCATTTGCTCAT 2112
QY 2041 TATGCTCAAAATGTTTCTATGTGGATATTTGCAATTTGCTGCTGCTTATTCATGATGTTT 2100
Db 2113 TATGCTCAAAATGTTTCTATGTGGATATTTGCAATTTGCTGCTGCTTATTCATGATGTTT 2172
QY 2101 GCTCTGTTGATATGCTGCTGAAATGCTGCAATGCTGCAATGCTGCTGCTGCTGCTGCTGCT 2160
Db 2173 GCTCTGTTGATATGCTGCTGAAATGCTGCAATGCTGCAATGCTGCTGCTGCTGCTGCTGCT 2232
QY 2161 CAGTGGGGTATTTCTTTTACAGAAATGCTGGGATGCTTTTGGGTTTGGGTTTGGGTTTGG 2220
Db 2233 CAGTGGGGTATTTCTTTTACAGAAATGCTGGGATGCTTTTGGGTTTGGGTTTGGGTTTGG 2292
QY 2221 CTTTATTTTCAATATTTGAACATAAAATCGTGTTCCTTATATAAATTTCTAG 2268
Db 2293 CTTTATTTTCAATATTTGAACATAAAATCGTGTTCCTTATATAAATTTCTAG 2340

KESULT 4
ABV25747
XX
AC
ARV25747;
NT
DE
DE
KW
KW
XX
OS
XX
PN
XX
PD
XX
PF
XX
PR
XX
PR
XX
PR
XX
PR
XX
```

```
PR 09-JUN-2000; 2000US-211314P.
PR 18-JUL-2000; 2000US-219007P.
PR 13-DEC-2000; 2000US-255281P.
XX
PA (MILL-) MILLENNIUM PREDICTIVE MEDICINE INC.
PI Schlegel R, Endege WO, Monahan JE;
XX
XX WPI; 2001-662795/76.
XX
PT Novel isolated nucleic acid molecule associated with cancerous state of
PT prostate cells and correlating with presence of prostate cancer, useful
PT for detecting presence of prostate cancer, stage of prostate cancer -
XX
PS Claim 1; Page 5169-5170; 11750pp; English.
XX
CC The invention relates to an isolated nucleic acid molecule (1) comprising
CC a nucleotide sequence given in Tables 1-9 (ABV00010-ABV62123) of the
CC specification or its complement. (1) is useful for:
CC (a) assessing whether a patient is afflicted with prostate cancer;
CC (b) monitoring the progression of prostate cancer in a patient;
CC (c) assessing the efficacy of a test compound to inhibit prostate
CC cancer in a patient;
CC (d) assessing the efficacy of a therapy for inhibiting prostate cancer
CC in a patient;
CC (e) selecting a composition for inhibiting prostate cancer in a patient;
CC (f) assessing the prostate cell carcinogenic potential of a compound;
CC (g) determining whether prostate cancer has metastasized in a patient;
CC (h) assessing the aggressiveness or indolence of prostate cancer in a
CC patient;
CC (1) is also useful as a pharmacodynamic or pharmacogenomic marker.
XX
SQ Sequence 3523 BP; 1071 A; 684 C; 735 G; 1013 T; 20 other;
Query Match 97.9%; Score 2220.4; DB 23; Length 3523;
Best Local Similarity 99.2%; Pred. No. 0;
Matches 2249; Conservative 0; Mismatches 1; Indels 18; Gaps 1;
QY 1 ATGGCGAGGAAGTTATCTGTAATCTTGATCCTGACCTTTGCCCTCTCTGTGCACAAATCCC 60
Db 138 ATGGCGAGGAAGTTATCTGTAATCTTGATCCTGACCTTTGCCCTCTCTGTGCACAAATCCC 197
QY 61 CTTCATGAACATAAAGCAGCTGCTTCCGCCAGACCACTGAGAAAATTTAGTCCGAATTTGG 120
Db 198 CTTCATGAACATAAAGCAGCTGCTTCCGCCAGACCACTGAGAAAATTTAGTCCGAATTTGG 257
QY 121 GAATCTGGCATTAATGTTGACTTTGGCAATTTCCACAGGCAATATCATCTACACAGCTT 180
Db 258 GAATCTGGCATTAATGTTGACTTTGGCAATTTCCACAGGCAATATCATCTACACAGCTT 317
QY 181 TTCTACCGCTATGGAGAAAATAATTTCTTGTACGTTGAAGGGTTTCAGAAAATTTACTTTCAA 240
Db 318 TTCTACCGCTATGGAGAAAATAATTTCTTGTACGTTGAAGGGTTTCAGAAAATTTACTTTCAA 377
QY 241 AATATAGCATAGATAAGATTTAAAGAAATCCATATACACCATGACCACTGACCACTCACTCA 300
Db 378 AATATAGCATAGATAAGATTTAAAGAAATCCATATACACCATGACCACTGACCACTCACTCA 437
QY 301 GACCAGGAGCATCACTCAGACCATGAGCGTCACTCAGACCATGAGCATCACTCAGACCATC 360
Db 438 GACCAGGAGCATCACTCAGACCATGAGCGTCACTCAGACCATGAGCATCACTCAGACCATC 497
QY 361 GAGCATCACTGACCATGATCATCTCTCACCATAATCATGCTGCTTCTGGTAAAT 420
Db 498 GAGCATCACTG-----ACCATAATCATGCTGCTTCTGGTAAAT 539
QY 421 AAGCGAAAAGCTCTTTGCCAGACCATGACTCAGATAGTTTCAGGTAAGATTCCTAGAAAC 480
Db 540 AAGCGAAAAGCTCTTTGCCAGACCATGACTCAGATAGTTTCAGGTAAGATTCCTAGAAAC 599
QY 481 AGCCAGGGGAAGAGCTCACCAGCAGCAACATCCCAAGTGGTGAAGGAATGTCAAGGAC 540
Db 600 AGCCAGGGGAAGAGCTCACCAGCAGCAACATCCCAAGTGGTGAAGGAATGTCAAGGAC 659
```

QY 541 ACTGTTACTGCTAGTGAAGTCACCTCAACTGCTGTACACACTGTCCTGGAAGCAACTCAC 600
DB 660 ACTGTTAGTGTAGTGAAGTCACCTCAACTGCTGTACACACTGTCCTGGAAGCAACTCAC 719
QY 601 TTTCCTAGAGACAATAGAGACTCCAAAGACCTTGGAAACCTCTTCCCAAGAGATCTTAACAGC 660
DB 720 TTTCCTAGAGACAATAGAGACTCCAAAGACCTTGGAAACCTCTTCCCAAGAGATCTTAAGCAGC 779
QY 661 TCCACTTCCACCACCTGTACACATCAAGAGCGGGGTGAGCGGCTGGCTGGTAGGAAACA 720
DB 780 TCCACTTCCACCACCTGTACACATCAAGAGCGGGGTGAGCGGCTGGCTGGTAGGAAACA 839
QY 721 AATGAATCTGTAGTGAGCCCGGAAAGGCTTTATGTATTCCAGAAACACAAATGAAAT 780
DB 840 AATGAATCTGTAGTGAGCCCGGAAAGGCTTTATGTATTCCAGAAACACAAATGAAAT 899
QY 781 CCTCAGAGTGTTCATATGCATCAAAAGCTTACGTACATCTCATGGCATGGGCATCCAGTT 840
DB 900 CCTCAGAGTGTTCATATGCATCAAAAGCTTACGTACATCTCATGGCATGGGCATCCAGTT 959
QY 841 CCGCTGAATGCAACAGAGTTCAACTATCTCTGTCCAGCCATCATCAACCAAATTTGATGCT 900
DB 960 CCGCTGAATGCAACAGAGTTCAACTATCTCTGTCCAGCCATCATCAACCAAATTTGATGCT 1019
QY 901 AGATCTTGTCTGATTCATACAAGTGAAGAGAGGCTGAAATCCCTCCAAAGACCTATTCA 960
DB 1020 AGATCTTGTCTGATTCATACAAGTGAAGAGAGGCTGAAATCCCTCCAAAGACCTATTCA 1079
QY 961 TTCAAAATAGCCGCTGGTGGTGTATATAGCCATTTCCATCATCAGTTTCCTGCTCTG 1020
DB 1080 TTCAAAATAGCCGCTGGTGGTGTATATAGCCATTTCCATCATCAGTTTCCTGCTCTG 1139
QY 1021 TCGGGGCTATCTTAGTGGCTCTCATGAATCGGGTGTATTTTCAAAATTTCTCTGACTTTC 1080
DB 1140 CTGGGGCTATCTTAGTGGCTCTCATGAATCGGGTGTATTTTCAAAATTTCTCTGAGTTTC 1199
QY 1081 CTCTGGCACTGGCGTTGGGACTTTGAGTGGTGATGCTTTTTTACACCTTCTTCCACAT 1140
DB 1200 CTCTGGCACTGGCGTTGGGACTTTGAGTGGTGATGCTTTTTTACACCTTCTTCCACAT 1259
QY 1141 TCTCATGCAAGTCACCACTATGTCATAGCCATGAAGAACCCAGCAATGGAATGCAAGA 1200
DB 1260 TCTCATGCAAGTCACCACTATGTCATAGCCATGAAGAACCCAGCAATGGAATGCAAGA 1319
QY 1201 GGACCACTTTTCAGTCATCTGCTCTCTCAAAACATAGAAGAGTGCCTATTTTCATTC 1260
DB 1320 GGACCACTTTTCAGTCATCTGCTCTCTCAAAACATAGAAGAGTGCCTATTTTCATTC 1379
QY 1261 ACCTGGAAGGGCTTAACAGCTCTAGGAGGCTGTATTTTCATGTTTCTGTTCAACATGTC 1320
DB 1380 ACCTGGAAGGGCTTAACAGCTCTAGGAGGCTGTATTTTCATGTTTCTGTTCAACATGTC 1439
QY 1321 CTCACATTCATCAACAAATTTAAAGATAAGAGAAAAGAAATCAGAGAACTCAAAAT 1380
DB 1440 CTCACATTCATCAACAAATTTAAAGATAAGAGAAAAGAAATCAGAGAACTCAAAAT 1499
QY 1381 CATGATGATGTGAGATTAAGAGAGAGTGTCCAAAGTATGAATCTCAACTTTCAACAAAT 1440
DB 1500 CATGATGATGTGAGATTAAGAGAGAGTGTCCAAAGTATGAATCTCAACTTTCAACAAAT 1559
QY 1441 GAGCAAGAGTACATACAGATGATCCAACTGAAGGCTATTTACGACGACACTCACAAGAG 1500
DB 1560 GAGCAAGAGTACATACAGATGATCCAACTGAAGGCTATTTACGACGACACTCACAAGAG 1619
QY 1501 CCCTCCCACTTTGATTTCTCAGCAGGCTGCGAGTCTTGGAAAGAGAGAGTCAATGATGCT 1560
DB 1620 CCCTCCCACTTTGATTTCTCAGCAGGCTGCGAGTCTTGGAAAGAGAGAGTCAATGATGCT 1679
QY 1561 CATGCTCATCCAGAGAACTCTACAATGAATATGTACCCAGAGGGTGGCAAGAAATAATGTC 1620
DB 1680 CATGCTCATCCAGAGAACTCTACAATGAATATGTACCCAGAGGGTGGCAAGAAATAATGTC 1739

QY 1621 CATTCACATTTCCACGATACACTCGGCCAGTCAGACGATCTCATTCACCACCATCATGAC 1680
DB 1740 CATTCACATTTCCACGATACACTCGGCCAGTCAGACGATCTCATTCACCACCATCATGAC 1799
QY 1681 TACCATCATATTCCTCCATCATCACCAACCAACCAACCAACCAACCAACCAACCAACCAAC 1740
DB 1800 TACCATCATATTCCTCCATCATCACCAACCAACCAACCAACCAACCAACCAACCAACCAAC 1859
QY 1741 CGCTACTCTCGGAGGAGCTGAAAGATGCGGCGTCCGCACCTTTGGCCTGGATGGTGATA 1800
DB 1860 CGCTACTCTCGGAGGAGCTGAAAGATGCGGCGTCCGCACCTCTGCGCTGGATGGTGATA 1919
QY 1801 ATGGGTGATGGGCTGCACAAATTTACCGGATGGCTAGCAAAATTTGGTGTCTCTTTACTGAA 1860
DB 1920 ATGGGTGATGGGCTGCACAAATTTACCGGATGGCTAGCAAAATTTGGTGTCTCTTTACTGAA 1979
QY 1861 GGCTTATCAAGTGGTTTAAAGTACTTCTGTGTGTCTGTCTGTCTGTCTGTCTGTCTGTCTGT 1920
DB 1980 GGCTTATCAAGTGGTTTAAAGTACTTCTGTGTGTCTGTCTGTCTGTCTGTCTGTCTGTCTGT 2039
QY 1921 TTAGTGACTTTGCTGTCTTACTAAAGGCTGGCATGACCGTTTAAAGCAGGCTCTCCTTTAT 1980
DB 2040 TTAGTGACTTTGCTGTCTTACTAAAGGCTGGCATGACCGTTTAAAGCAGGCTCTCCTTTAT 2099
QY 1981 AATGCAATTTGTCAGCCATGCTGGGTATCTTGGAAATGGCAACAGAGAAATTTTCATTTGTCAT 2040
DB 2100 AATGCAATTTGTCAGCCATGCTGGGTATCTTGGAAATGGCAACAGAGAAATTTTCATTTGTCAT 2159
QY 2041 TATGCTGAAATTTTCTTATGTGGATATTTGCACCTTACTGCTGGCTTATTCATGTATGTT 2100
DB 2160 TATGCTGAAATTTTCTTATGTGGATATTTGCACCTTACTGCTGGCTTATTCATGTATGTT 2219
QY 2101 GCTCTGGTTGATATGGTACCTGAAATGCTGCACAAATGATGCTAGTGACCATGGATGTAGC 2160
DB 2220 GCTCTGGTTGATATGGTACCTGAAATGCTGCACAAATGATGCTAGTGACCATGGATGTAGC 2279
QY 2161 CGCTGGGGTATTTCTTTTACAGAAATGCTGGGATGCTTTTGGGTTTGGGAATTTATGTTA 2220
DB 2280 CGCTGGGGTATTTCTTTTACAGAAATGCTGGGATGCTTTTGGGTTTGGGAATTTATGTTA 2339
QY 2221 CTTATTTCCATATTTGACATAAAATCGTGTTCGTATATAATTTCTAG 2268
DB 2340 CTTATTTCCATATTTGACATAAAATCGTGTTCGTATATAATTTCTAG 2387

RESULT 5
AAD13479 standard; DNA; 3461 BP.
XX AAD13479;
AC AAD13479;
XX
DT 06-NOV-2001 (first entry)
XX Human LIV-1 gene.
DE Human; LIV-1; cytostatic; estrogen-inducible gene; tumour; therapy;
XX cancer; breast; lung; prostate; colon; ovary; uterus; kidney; gastric;
KW salivary gland; carcinoma; drug screening; ds.
XX Homo sapiens.
XX Key Location/Qualifiers
FH 138..2396
CDS /*tag= a
FT /product= "Human LIV-1 protein"
FT 138..197
FT /*tag= b
FT 198..2393
FT mat_peptide /*tag= c
FT /product= "Mature LIV-1 protein"
XX
PN W0200155178-A2.
XX

```
PD 02-AUG-2001.
XX 25-JAN-2001: 2001WO-US02622.
XX 25-JAN-2000: 2000US-0177951.
XX 10-APR-2000: 2000US-0195761.
XX (GENE) GENENTECH INC.
XX
XX Goddard A, Gurney AL, Smith V, Hongo JS, De Sauvage F;
XX WPL: 2001-502628/55.
XX P-PSDB: AA017205.
XX
XX New LIV-1 DNA and amino acids, useful for treating cancer or inhibiting
XX the proliferation of tumor cells in mammals, e.g. breast, lung,
XX prostate, colon, ovary, uterus, kidney, gastric or salivary carcinoma
XX .
XX
XX Example 2; Fig 1A; 150pp; English.
XX
XX The present sequence is human estrogen-inducible gene, (LIV-1).
XX LIV-1 is overexpressed in tumour tissues such as prostate, colon,
XX lung and breast. The LIV-1 DNA and polypeptide, and the monoclonal
XX antibody are useful for treating cancer and inhibiting the proliferation
XX of tumour cells in mammals, e.g. breast, lung, prostate, colon, ovary,
XX uterus, kidney, gastric or salivary carcinoma, or other tumour cell
XX types expressing the LIV-1-164647 protein. In particular, the mammal is
XX a human. The LIV-1 DNA and polypeptide may also be used in screening
XX assays for drug candidates.
XX
XX Sequence 3461 BP; 1057 A; 679 C; 727 G; 998 T; 0 other;
XX
Query Match          96.9%; Score 2198; DB 22; Length 3461;
Best local similarity 99.1%; Prod. No. 0;
Matches 2248; Conservative 0; Mismatches 0; Indels 20; Gaps 3;
XX
QY 1 ATGGCGAGCAAGTATCTGTATCTTGATCTGACCTTGGCCCTCTCTGTCTCAAAATGCC 60
DB 1 ATGGCGAGCAAGTATCTGTATCTTGATCTGACCTTGGCCCTCTCTGTCTCAAAATGCC 197
XX
QY 61 CTTCATGCAAGTAAAGCAGCTCTTTTCCCGCAGACCTTGAGAAATTTAGTCCGAATTCG 120
DB 1 CTTCATGCAAGTAAAGCAGCTCTTTTCCCGCAGACCTTGAGAAATTTAGTCCGAATTCG 257
XX
QY 121 GAATCTGCGATTAATGTTGACCTTGCAATTTCCACAGCGCAATATCATCTACACAGCTT 180
DB 1 GAATCTGCGATTAATGTTGACCTTGCAATTTCCACAGCGCAATATCATCTACACAGCTT 317
XX
QY 181 TTCTACCTGATGAGAAATTAATCTTGTGTCAGTTGAAGGTTTCAGAAATTTACTTCAA 240
DB 1 TTCTACCTGATGAGAAATTAATCTTGTGTCAGTTGAAGGTTTCAGAAATTTACTTCAA 377
XX
QY 241 AATATAGGCATAGATAAGATTTAAAGAAATCCATATACACCATGACGACGACATCACTCA 300
DB 1 AATATAGGCATAGATAAGATTTAAAGAAATCCATATACACCATGACGACGACATCACTCA 437
XX
QY 301 GACCCAGGACATCACTAGACCATGAGCGTCACTCAGACCATGAGCATCACTCAGACCAT 360
DB 1 GACCCAGGACATCACTAGACCATGAGCGTCACTCAGACCATGAGCATCACTCAGACCAT 497
XX
QY 361 GAGCATCACTCTGACCATGATCATCACTCTCCACCATATCATCTGCTTCTGTTAAAT 420
DB 1 GAGCATCACTCTGACCATGATCATCACTCTCCACCATATCATCTGCTTCTGTTAAAT 539
XX
QY 421 AAGTGAAAGCTCTTTGCCCCAGACCATGACTCAGATAGTTTCAGGTAAAGATCTTACAAAC 480
DB 1 AAGTGAAAGCTCTTTGCCCCAGACCATGACTCAGATAGTTTCAGGTAAAGATCTTACAAAC 539
XX
QY 481 ACCGAGGGAAGAGCTTCCCGACCGACGACATCCCACTGGTACAGGAATTTTAAAGAC 540
DB 1 ACCGAGGGAAGAGCTTCCCGACCGACGACATCCCACTGGTACAGGAATTTTAAAGAC 659
XX
QY 541 AGTCTTAGTCTAGTGAAGTCACTCACTCACTGTGTACAAACATCTGTCTCTGAAGGAACTCAC 600
DB 1 AGTCTTAGTCTAGTGAAGTCACTCACTCACTGTGTACAAACATCTGTCTCTGAAGGAACTCAC 719
XX
QY 601 TTTCTAGAGACAATAGAGACTTCAAGACCTTCCAAAACCTTTTCCCAAGAGATGTAAAGCAGC 660
DB 1 TTTCTAGAGACAATAGAGACTTCAAGACCTTCCAAAACCTTTTCCCAAGAGATGTAAAGCAGC 779
XX
QY 661 TCCACTCCACCCAGTGTGCATCAATCAAGAGCGCGGTGAGCCGGCTGGTGTAGGAAACA 720
DB 1 TCCACTCCACCCAGTGTGCATCAATCAAGAGCGCGGTGAGCCGGCTGGTGTAGGAAACA 839
XX
QY 721 AATGAATCTGTGAGTGAAGCCCGGAAAGGCTTTTATGTATTTCCAGAAACACAAATGAAAT 780
DB 1 AATGAATCTGTGAGTGAAGCCCGGAAAGGCTTTTATGTATTTCCAGAAACACAAATGAAAT 899
XX
QY 781 CCTCAGGAGTGTTCATATGATCAATCAAGCTACTGACATCTCATGCGATGGGATCCAGGTT 840
DB 1 CCTCAGGAGTGTTCATATGATCAATCAAGCTACTGACATCTCATGCGATGGGATCCAGGTT 959
XX
QY 841 CCGCTGAATGCAACAGAGTTCAACTATCTGTGTCCAGCCATCATCAACCAATTTGATGCT 900
DB 1 CCGCTGAATGCAACAGAGTTCAACTATCTGTGTCCAGCCATCATCAACCAATTTGATGCT 1019
XX
QY 901 AGATCTTGTCTGTATCATCAAGTGAAGAGGCTCAAAATCCCTCCAAAGACCTATTCA 960
DB 1 AGATCTTGTCTGTATCATCAAGTGAAGAGGCTCAAAATCCCTCCAAAGACCTATTCA 1079
XX
QY 961 TTACAAATAGCCTGGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGG 1020
DB 1 TTACAAATAGCCTGGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGG 1139
XX
QY 1021 CTGGGGCTTATCTTAGTGCCTCTCATCAATCGGGGTGGTGGTGGTGGTGGTGGTGGTGGTGG 1080
DB 1 CTGGGGCTTATCTTAGTGCCTCTCATCAATCGGGGTGGTGGTGGTGGTGGTGGTGGTGGTGG 1199
XX
QY 1081 CTTGTGGCAGCTGGCGGTGGGACTTTTGCAGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGG 1140
DB 1 CTTGTGGCAGCTGGCGGTGGGACTTTTGCAGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGG 1259
XX
QY 1141 TCTCATGCAAGTACACACCATAGTGTATAGCCATGAAGAACCAAGCAATGGAATGAAAGA 1200
DB 1 TCTCATGCAAGTACACACCATAGTGTATAGCCATGAAGAACCAAGCAATGGAATGAAAGA 1319
XX
QY 1201 GGACCACTTTTCACTCATCTCTCTCAAAACATAGAAAGAGTGCCTATTGTTGATTC 1260
DB 1 GGACCACTTTTCACTCATCTCTCTCAAAACATAGAAAGAGTGCCTATTGTTGATTC 1379
XX
QY 1261 ACCTGGAAGGCTCTAACAGCTCTAGGAGGCTGTATTTTCTATGTTTCTTGAACATGTC 1320
DB 1 ACCTGGAAGGCTCTAACAGCTCTAGGAGGCTGTATTTTCTATGTTTCTTGAACATGTC 1439
XX
QY 1321 CTCACTTTGATCAACAAATTTAAAGATAGAAAGAAAGAAATCAGAGAAACCTGAAAT 1380
DB 1 CTCACTTTGATCAACAAATTTAAAGATAGAAAGAAAGAAATCAGAGAAACCTGAAAT 1499
XX
QY 1381 GATGATGATGGAGATTAAGAAGCAGTGTCTCCCAAGTATGAATCTCAACTTTTCAACAAAT 1440
DB 1 GATGATGATGGAGATTAAGAAGCAGTGTCTCCCAAGTATGAATCTCAACTTTTCAACAAAT 1559
XX
QY 1441 GAGGAGAAAGTATAGATACAGATGATCGTGAAGGCTATTATTCGAGCAGACATCACAAGAG 1500
DB 1 GAGGAGAAAGTATAGATACAGATGATCGTGAAGGCTATTATTCGAGCAGACATCACAAGAG 1619
XX
QY 1501 CCTCTCCCACTTTGATCTCAGCAGCCTGAGTCTTGAAGAGAGAGAGTCTATGATAGCT 1560
DB 1 CCTCTCCCACTTTGATCTCAGCAGCCTGAGTCTTGAAGAGAGAGAGTCTATGATAGCT 1679
XX
QY 1561 CATGCTCATCCACGAGAGTCTACAATGAATATGTACCCAGAGGGTGCAGAAATTAATGTC 1620
DB 1 CATGCTCATCCACGAGAGTCTACAATGAATATGTACCCAGAGGGTGCAGAAATTAATGTC 1739
XX
QY 1621 CATTCACATTTTCCACGATACACTCGGCCAGTCAGAGCATCTCTATTACACCACCATCATGAC 1680
```

db	1740	CATTTCACATTTCCACGATACACTCGGCCAGTTCAGAGGATCTCATTTCCACACGATCATGAC	1799
QY	1681	TACCATCATATTTCCATCATCAACCACCACCAAAACACCATCTCTCACAGTTCACAGCCAG	1740
Db	1800	TACCATCATATTTCTCCATCATCAACCACCACCAAAACACCATCTCTCACAGTTCACAGCCAG	1859
QY	1741	CCTACACTCTCGGGAGGAGCTGAAGATGCGCGGCTGCGCACTTTGGCTTGGATGCTGATGA	1800
Db	1860	CGCTACTCTCGGGAGGAGCTGAAGATGCGCGGCTGCGCACTTTGGCTTGGATGCTGATGA	1919
QY	1801	ATGGCTGATGCCCTGCGACAAATTTACAGCGATGGCCCTAGCAATTTGGTCTGCTTTTACTCAA	1860
Db	1920	ATGGCTGATGCCCTGCGACAAATTTACAGCGATGGCCCTAGCAATTTGGTCTGCTTTTACTGAA	1979
QY	1861	GGCTTATCAAGTGGTTTAAGTACTTCTGTGTGCTGTCTGTGCATGAGTTGCCATCATGAA	1920
Db	1980	GGCTTATCAAGTGGTTTAAGTACTTCTGTGTGCTGTCTGTGCATGAGTTGCCATCATGAA	2039
QY	1921	TTAGGTGACATTTTGTCTTCTACTTAAAGGCTGGCATGACCGTTTAAGCAGCGTGTCTTTAT	1980
Db	2040	TTAGGTGACATTTTGTCTTCTACTTAAAGGCTGGCATGACCGTTTAAGCAGCGTGTCTTTAT	2099
QY	1981	AATCCATTTGCAGCGATGCTGGCGTATCTTTGGAATGGCAACAGGAATTTTCATTTGCTCAT	2040
Db	2100	AATCCATTTGCAGCGATGCTGGCGTATCTTTGGAATGGCAACAGGAATTTTCATTTGCTCAT	2159
QY	2041	TATGCTCAAAATGTTTCTATGTGGATATTTCACATTTACTGCTGCTTATTTTCATGCTATT	2100
Db	2160	TATGCTCAAAATGTTTCTATGTGGATATTTCACATTTACTGCTGCTTATTTTCATGCTATT	2219
QY	2101	GCTCTGCTTGATATGTTTACCTCAAAATGCTGCACAATGATGCTAGTGACCATGATGTAGC	2160
Db	2220	GCTCTGCTTGATATGTTTACCTCAAAATGCTGCACAATGATGCTAGTGACCATGATGTAGC	2279
QY	2161	CGCTGGGGGTATTTCTTTTACAGAATGCTGGGATGCTTTTGGGTTTGGAAATATGTTA	2220
Db	2280	CGCTGGGGGTATTTCTTTTACAGAATGCTGGGATGCTTTTGGGTTTGGAAATATGTTA	2339
QY	2221	CTTATTTTCATATTTGAACATAAAATCGTGTTCGTATAAAATTTCTAG	2268
Db	2440	CTTA-TTCCATATTTCAACATAAAATCGTG-TTCGTATAAAATTTCTAG	2385
RESULT 6			
10	AAT99070		
XX	AAT99070	standard; cdna; 1310 BP.	
AC	AAT99070;		
XX			
DT	24-MAR-1998	(first entry)	
XX			
DE		Partial sequence of the pLIV1 gene.	
XX			
KW		Oestrogen regulated gene; pLIV1 gene partial sequence; breast cancer;	
KW		metastatic risk assessment; hormone therapy responsiveness;	
KW		lumour oestrogen receptor expression; lymph node involvement; ds.	
XX			
OS		homo sapiens.	
XX			
FH		Key	
FT		Location/Qualifiers	
FT		1..1296	
XX		/*tag= a	
PN		US5693465-A.	
XX			
PD		02-DEC-1997.	
XX			
PF		22-SEP-1994; 94US-0311023.	
XX			
PR		22-SEP-1994; 94US-0311023.	
XX			
PA		(UYWA-) UNIV WALES COLLEGE OF MEDICINE.	

Gee JMW, Green CD, Manning DL, Nicholson RI;
 WPI: 1998-031740/03.
 P-PSDB: AAW34528.
 Assessment of metastatic risk or oestrogen responsive-ness in breast
 cancer - by detecting expression of pLlVI protein
 Claim 20; Column 5-10; 11pp; English.
 This sequence represents a partial sequence of the plvlI gene. The
 protein encoded by this sequence is an oestrogen regulated gene
 associated with breast cancer. The protein encoded by this sequence can
 be used in the method of the invention. The method is for determining the
 risk of metastasis of a female breast tumour and predicting the
 responsiveness of a female breast tumour to hormone treatment comprise
 determining if a tissue sample from the tumour expresses a polypeptide
 comprising at least 14 continuous amino acids of the protein encoded by
 this sequence. The method is useful for determining the risk of
 metastasis of a female breast tumour and to predict the responsive of
 female breast tumour to hormone treatment. plvlI expression is highly
 predictive of tumour oestrogen receptor expression and lymph node
 involvement.
 Sequence 1310 BP: 354 A; 271 C; 287 G; 398 T; 0 other;

Query Match	55.7%	Score 1263	DB 19	Length 1310
Best Local Similarity	99.8%	Pred. No. 0		
Matches 1285	Conservative 0	Mismatches 0	Indels 2	Gaps 2
QY	982	GGTTTATAGCCATTTCCATCATCAGTTTCTGTCTCTGCTGGGGTTCCTTAGTGCTT	1041	
Db	1	GGTTTATAGCCATTTCCATCATCAGTTTCTGTCTCTGCTGGGGTTCCTTAGTGCTT	60	
QY	1042	CTCATGAATCGGGTGTTTTTCCAAATTTCTCTGAGTTTCTTGTCGACATGCGCCGTGGG	1101	
Db	61	CTCATGAATCGGGTGTTTTTCCAAATTTCTCTGAGTTTCTTGTCGACATGCGCCGTGGG	120	
QY	1102	ACTTTGAGTGGTGATGCTTTTTTACACCTTCTTCCACATTCCTCATGCAAGTCACCA	1161	
Db	121	ACTTTGAGTGGTGATGCTTTTTTACACCTTCTTCCACATTCCTCATGCAAGTCACCA	180	
QY	1162	AGTCATAGCCCATGAAGAACAGCAATGAAATGAAACAGGACCACTTTTCAGTCATC	1221	
Db	181	AGTCATAGCCCATGAAGAACAGCAATGAAATGAAACAGGACCACTTTTCAGTCATC	240	
QY	1222	TCTTCTCAAAACATAGAGAAAGTGCCCTATTTTGATTTCCAGTGTGGAAGGGTCTAAC	1281	
Db	241	TCTTCTCAAAACATAGAGAAAGTGCCCTATTTTGATTTCCAGTGTGGAAGGGTCTAAC	300	
QY	1282	CTAGGAGGCGCTGATTTTCATGCTTTCTTGTCGAACATGTCCCTACATTCATGATCA	1341	
Db	301	CTAGGAGGCGCTGATTTTCATGCTTTCTTGTCGAACATGTCCCTACATTCATGATCA	360	
QY	1342	AAAGATAAGAGAAAAAGAAATCAGAAGAAACCTTGAAATGATGATGTGCGAGATTAA	1401	
Db	361	AAAGATAAGAGAAAAAGAAATCAGAAGAAACCTTGAAATGATGATGTGCGAGATTAA	420	
QY	1402	AAGCAGTTGTCCAGTATGAATCTCAACTTTTCAACAAATGAGGAAAGTAGATACAGAT	1461	
Db	421	AAGCAGTTGTCCAGTATGAATCTCAACTTTTCAACAAATGAGGAAAGTAGATACAGAT	480	
QY	1462	GATCGAACTGAGGCTATTTACGAGCAGACATCAAGAGCCCTCCCACCTTTGATTC	1521	
Db	481	GATCGAACTGAGGCTATTTACGAGCAGACATCAAGAGCCCTCCCACCTTTGATTC	540	
QY	1522	CAGCCTCGAGTCTTGGGAAGAGAGAGTCTATGATAGCTCATGCTCATCCACAGGA	1581	
Db	541	CAGCCTCGAGTCTTGGGAAGAGAGAGTCTATGATAGCTCATGCTCATCCACAGGA	600	
QY	1582	TACAACTAATATGTACCAGAGGGTGCAAGAAATAAATGCCATTCACATTTCCAGTA	1641	

RESULT#	8
AAT33219	
ID	AAT33219 standard; cDNA; 1310 BP.
XX	XX
XX	AAT33219;
XX	XX
DT	16-OCT-1996 (first entry)
XX	XX
DE	Oestrogen-regulated pLlV1 gene partial cDNA clone.
XX	XX
KW	Oestrogen-regulated gene; pLlV1; breast cancer; tumour; diagnosis;
KW	market; metastasis; ss.
XX	XX
OS	Homo sapiens.
XX	XX
FH	Location/Qualifiers
FT	1..1296
CDS	/tag= a
FT	XX
XX	XX
PN	CA2132500-A.
XX	XX
PD	21-MAR-1996.

Ub	541	CAGCCTCGACCTCTTGGGAAGAGAGGTGATGATACCTTCATGCTCAGAGAAATC	600
Qy	1582	TATCAATCAATATCTACCCAGAGGGTGCAGAATAAATGCCATTCACATTTCTACGATACA	1641
Db	601	TACAAATCAATATCTACCCAGAGGGTGCAGAATAAATGCCATTCACATTTCTACGATACA	660
Qy	1642	CTCGGGCAGTTCAGAGGATCTCATTTACCAACCAATCATGACTAGCAATCATATTTCTGCATCAT	1701
Db	661	CTCGGGCAATTCAGAGGATCTCATTTACCAACCAATCATGACTAGCAATCATATTTCTGCATCAT	720
Qy	1702	CACCACCACCAAAACCACTCTCACAGTCACGCCAGCGCTACTCTCGGAGAGACCTG	1761
Db	721	CACCACCACCAAAACCACTCTCACAGTCACGCCAGCGCTACTCTCGGAGAGAGCTG	780
Qy	1762	AAAGATCGCGCGCTCGCCACTTTGGCGCTGGATGGTGAATAATGGGTGATGGCTGCGACAAT	1821
Ub	781	AAAGATCGCGCGCTCGCCACTTTGGCGCTGGATGGTGAATAATGGGTGATGGCTGCGACAAT	840
Qy	1822	TTACAGCATGGCGTTAGCAATTTGGTGGTGTCTTTTACTGAAGGCTTTATCAAGTGGTTTAAAT	1881
Db	841	TTACAGCATGGCGTTAGCAATTTGGTGGTGTCTTTTACTGAAGGCTTTATCAAGTGGTTTAAAT	900
Qy	1882	ACTTCTCTTGGCTGTCTGTCTCATGAGTTGGCCCTCATGAATTAGGTGACTTTGCTGTCTTA	1941
Db	901	ACTTCTCTTGGCTGTCTGTCTCATGAGTTGGCCCTCATGAATTAGGTGACTTTGCTGTCTTA	960
Qy	1942	CTAAAGCGTCGCATGATCCGTTAAGCAGGCTGTCTTTTATAATGCATCTGCAGCCATGCTG	2001
Db	961	CTAAAGCGTCGCATGATCCGTTAAGCAGGCTGTCTTTTATAATGCATCTGCAGCCATGCTG	1020
Qy	2002	GCCTATCTTGGAAATGGCAACAGGAATTTTCATTGGTGCATTTAGCTGAAATCTTTCTATG	2061
Db	1021	GCCTATCTTGGAAATGGCAACAGGAATTTTCATTGGTGCATTTAGCTGAAATCTTTCTATG	1080
Qy	2062	TGATATTTTGCACCTTACTGCTGCCCTTATTCATGATGTGTTGCTCTGGTGTGATATGGTACCT	2121
Db	1081	TGGATATTTGCACCTTACTGCTGCCCTTATTCATGATGTGTTGCTCTGGTGTGATATGGTACCT	1140
Qy	2122	GAATGCTGCACAATCATGCTACTGACCATGCATGACCTAGCCGCTGGGGGATTTCTTTTAA	2181
Db	1141	GAATGCTGCACAATCATGCTACTGACCATGCATGACCTAGCCGCTGGGGGATTTCTTTTAA	1200
Qy	2182	CACAATGCTGGGATGCTTTTGGGTTTGGGAATATGTTTACTTATTTTCCATATTTTGAACAT	2241
Ub	1201	CACAATGCTGGGATGCTTTTGGGTTTGGGTTTGGGAATATGTTTACTTATTTTCCATATTTTGAACAT	1259
Qy	2242	AAATCTGTGTTTCGTTATAAATTTCTAG	2268
Ub	1260	AAATCTGTGTTTCGTTATAAATTTCTAG	1285

RESULT 9	
AAT33220	
11)	AAT33220 standard; cDNA; 2404 BP.
XX	
XX	
AC	AAT33220;
XX	
XX	
DT	16-OCT-1996 (first entry)
XX	
DE	Oestrogen-regulated p11v1 gene partial 3' extended cDNA clone.
XX	
XX	Oestrogen-regulated gene; p11v1; breast cancer; tumour; diagnosis;
KW	marker; metastasis; ss.
KW	

OS	Location/Qualifiers
XX	
FH	key
FT	1..1296
FT	CDS
FT	/*tag= a
FT	3' UTR
FT	1297..2404
FT	/*tag= b
FT	misc_difference
FT	2201..2260
FT	/*tag= c

FT	/note= "bases 2201-2260 are identical to bases 2141-2200"
FT	
XX	
PN	CA2132500-A.
XX	
PD	21-MAR-1996.
XX	
PF	20-SEP-1994; 94CA-2132500.
XX	
PR	20-SEP-1994; 94CA-2132500.
XX	
PA	(UYWA-) UNIV WALES COLLEGE OF MEDICINE.
XX	
PI	Gee JM, Green CB, Manning DL, Nicholson RI;
DR	WPI; 1996-268987/28.
DR	P-PSDB; AAR98004.
XX	
PT	Oestrogen-regulated plivl gene and deduced polypeptide - useful for
PT	predicting the propensity for metastatic spread or the
PT	responsiveness to endocrine treatment of breast tumour
XX	
PS	Disclosure; Page 16-17; 27pp; English.
XX	
CC	A cDNA clone (AAT33219) corresponding to a portion of the oestrogen-
CC	regulated human plivl gene was extended (AAT33220) using a primer
CC	directed cloning strategy. plivl is a candidate for the hormonal
CC	regulation of tumour invasion of breast cancer, a prognostic marker
CC	of metastatic spread, and a suitable target for therapeutic
CC	intervention by antihormones and antibody-directed methods. It is
CC	significantly associated with oestrogen-dependent (ER+) disease.
CC	The polypeptide (AAR98004) encoded by the cDNA contains 3 immunogenic
XX	regions.
XX	
SQ	Sequence 2404 BP; 726 A; 413 C; 490 G; 775 T; 0 other;
<p>Query Match 55.5%; Score 1259.8; DB 17; Length 2404; Best Local Similarity 99.7%; Pred. No. 0; Matches 1283; Conservative 0; Mismatches 2; Indels 2; Gaps 2;</p>	
QY	982 GGTTTATTAGCCATTCCATCATCAGTTTCTGTCTGCTGGGGGTTATCTTAGTGCCT 1041
DB	1 GGTTTTATAGCCATTCCATCATCAGTTTCTGTCTGCTGGGGGTTATCTTAGTGCCT 60
QY	1042 CTCATGAATCGGGTGTTTTCAAATTTCTCTGAGTTTCTTGCGCACTGGCGTTGGG 1101
DB	61 CTCATGAATCGGGTGTTTTCAAATTTCTCTGAGTTTCTTGCGCACTGGCGTTGGG 120
QY	1102 ACTTTGAGTGTGATGCTTTTTTACACTTCTTCCACATTTCTCATGCAAGTCACCACT 1161
DB	121 ACTTTGAGTGTGATGCTTTTTTACACTTCTTCCACATTTCTCATGCAAGTCACCACT 180
QY	1162 AGTCATAGCCATGAAGAACAGCAATGGAATGAAAGAGGAGGACCACATTTTCAGTCTCTG 1221
DB	181 AGTCATAGCCATGAAGAACAGCAATGGAATGAAAGAGGAGGACCACATTTTCAGTCTCTG 240
QY	1222 TCTTCTCAAACATAGAGAAAGTGCCATTTTTCATTCCAGTCGGAAGGCTCTTAACAGCT 1281
DB	241 TCTTCTCAAACATAGAGAAAGTGCCATTTTTCATTCCAGTCGGAAGGCTCTTAACAGCT 300
QY	1282 CTAGGAGGCTGTATTTTCATGTTTCTTGTGAACATGCTCCTCACATTTGATCAACAATTT 1341
DB	301 CTAGGAGGCTGTATTTTCATGTTTCTTGTGAACATGCTCCTCACATTTGATCAACAATTT 360
QY	1342 AAAGATAGAGAAAAAGAAATCAGAGAAACCTGAAATGATGATGTGGAGATTAAAG 1401
DB	361 AAAGATAGAGAAAAAGAAATCAGAGAAACCTGAAATGATGATGTGGAGATTAAAG 420
QY	1402 AAGCAGTTGTCCAGTATGAATCTCAACTTTCACAAATGAGGAGAAAGTAGATACAGAT 1461
DB	421 AAGCAGTTGTCCAGTATGAATCTCAACTTTCACAAATGAGGAGAAAGTAGATACAGAT 480
QY	1462 GATCGAACTGAAGGCTATTTACGAGCAGACTCAACAGAGCCCTCCCACTTTGATTCTCAG 1521

```
db 481 GATCGAAGTCAAGGCTATTTACACAGACTCACAGAGCCCTCCACATTTGATTTCTCAG 540
1522 CAGCCTGCGAGTCTTGGAGAAAGAGGTCATGATGCTCATGCTCATCAAGAGAGTC 1581
541 CAGCCTGAGTCTTGGAGAAAGAGGTCATGATGCTCATGCTCATCAAGAGAGTC 600
1582 TACAATGAATATGTTACCCAGAGGGTGCAGAAATAAATGCCATTTCACATGATACA 1641
601 TACAATGAATATGTTACCCAGAGGGTGCAGAAATAAATGCCATTTCACATGATACA 660
1642 CTCGGCCAGTCAGACCATCTTACCCAGCATCATGATGCTCATGCTCATCAATAT 1701
661 CTCGGCCAGTCAGACCATCTTACCCAGCATCATGATGCTCATGCTCATCAATAT 720
1702 CACCACCAACCAACCATCTTACAGTCACAGCAGCGTACTCTCGGAGAGCTG 1761
721 CACCACCAACCAACCATCTTACAGTCACAGCAGCGTACTCTCGGAGAGCTG 780
1762 AAGATGCGCGGCTCGCCACTTTGGCCCTGGATGGTGATAATGGTGATGGCTGCACAA 1821
781 AAGATGCGCGGCTCGCCACTTTGGCCCTGGATGGTGATAATGGTGATGGCTGCACAA 840
1822 TTCAGCGATGCGCTAGCAATTTGGTCTTTTACTGAAGGCTTATCAAGTCTTTAAGT 1881
841 TTCAGCGATGCGCTAGCAATTTGGTCTTTTACTGAAGGCTTATCAAGTCTTTAAGT 900
1882 ACTTCGTTGCTGCTGCTGCTCATGATGCTGCTCATGATGCTGCTGCTGCTGCTA 1941
901 ACTTCGTTGCTGCTGCTGCTCATGATGCTGCTCATGATGCTGCTGCTGCTGCTA 960
1942 CTAAGGCTGGCATGACCTTAAGCAGGCTGCTCTTTTATAATGCATTTGACGCCATGCTG 2001
961 CTAAGGCTGGCATGACCTTAAGCAGGCTGCTCTTTTATAATGCATTTGACGCCATGCTG 1020
2002 GCGTATCTTGGAAATGCAACAGCAATTTTCATGCTGCTCATGCTGCTGCTGCTATG 2061
1021 GCGTATCTTGGAAATGCAACAGCAATTTTCATGCTGCTCATGCTGCTGCTGCTATG 1080
2062 TCGATATTTGCACTTACTGCTGCTGCTTATCATGATGCTGCTGCTGCTGCTGCTGCT 2121
1081 TCGATATTTGCACTTACTGCTGCTGCTTATCATGATGCTGCTGCTGCTGCTGCTGCT 1140
2122 GAAATGCTGCAATGATGCTAGTGACCATGATGCTAGCGGTGCTGCTGCTGCTGCTGCT 2181
1141 GAAATGCTGCAATGATGCTAGTGACCATGATGCTAGCGGTGCTGCTGCTGCTGCTGCT 1200
2182 CAGATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 2241
1201 CAGATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1259
2242 AAAATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 2268
1260 AAAATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1285

RESULT 10
ID AAF16156
XX AAF16156 standard; cdna; 1193 BP.
AC AAF16156;
XX
DT 13-MAR-2001 (first entry)
XX
DE human prostate cancer antigen nucleotide sequence SEQ ID NO:591.
XX
KW Human; prostate cancer; prostate cancer antigen; detection; diagnosis;
KW neuroprotective; cytoskeletal; cardioactive; immunomodulatory; muscular;
KW vulnary; gastrointestinal; nephrotropic; antiinfective; gynaecological;
KW antibacterial; gene therapy; neural; immune; reproductive; renal;
KW gastrointestinal; pulmonary; cardiovascular; proliferative disorder;
KW wound; infectious disease; ss.
XX
```

```
OS Homo sapiens.
PN WO20005174-A1.
PD 21-SEP-2000.
XX
PF 08-MAR-2000; 2000WO-US05988.
PR 12-MAR-1999; 99US-0124270.
XX
XX (HUMA-) HUMAN GENOME SCI INC.
XX (ROSE/) ROSEN C A.
XX Rosen CA, Ruben SM;
XX
XX WPI: 2000-587513/55.
XX P-PSDB: AAB56953.
XX
XX Prostate cancer associated gene sequences, referred to as prostate
XX disorders such as prostate cancer -
XX
XX Claim 1: Page 1045-1046; 2338pp; English.
XX
XX AAF15566 to AAF16505 encode the human prostate cancer associated
XX proteins, called prostate cancer antigens, given in AAB56363 to AAB57302.
XX The prostate cancer antigens can have neuroprotective, cytoskeletal,
XX cardioactive, immunomodulatory, muscular, vulnary, gastrointestinal,
XX nephrotropic, antiinfective, gynaecological and antibacterial activities,
XX and can be used in gene therapy. The prostate cancer antigen
XX polynucleotides may be used for detection of prostate cancer, chromosome
XX identification, as chromosome markers, and for numerous other diagnostic
XX or research purposes. The prostate cancer antigens may be used to treat
XX disorders such as neural, immune, muscular, reproductive,
XX gastrointestinal, pulmonary, cardiovascular, renal, and proliferative
XX disorders, wounds, and infectious diseases. AAF16506 to AAF16514 to
XX AAB57303 represent sequences used in the exemplification of the present
XX invention.
XX
XX Sequence 1193 BP; 393 A; 259 C; 235 G; 302 T; 4 other;
XX
Query Match 50.3%; Score 1141.4; DB 21; Length 1193;
Best Local Similarity 99.7%; Pred. No. 1.2e-301;
Matches 1139; Conservative 4; Mismatches 0; Indels 0; Gaps 0;
QY 539 ACAGTGTAGTCTAGTCAAGTGAACCTCAACTGTGTACACACTGTCTCTGAAGGAACCTC 598
DB 1 ACAGTGTAGTCTAGTCAAGTGAACCTCAACTGTGTACACACTGTCTCTGAAGGAACCTC 60
QY 599 ACTTCTTAGACACAATAGAGACTCCAAAGACTTGGAAACTTCTCCCAAGATCTAAGCA 658
DB 61 ACTTCTTAGACACAATAGAGACTCCAAAGACTTGGAAACTTCTCCCAAGATCTAAGCA 120
QY 659 GCTCCACTCCACCAGTGTACATCAAGAGCGGGTGAGCGGCTGGCTGGTAGGAAAA 718
DB 121 GCTCCACTCCACCAGTGTACATCAAGAGCGGGTGAGCGGCTGGCTGGTAGGAAAA 180
QY 719 CAATGAATCTGTGAGTGAAGCGGCGGAGGCGGCTTTATGTTATCCAGAAACACAAATGAAA 778
DB 181 CAATGAATCTGTGAGTGAAGCGGCGGAGGCGGCTTTATGTTATCCAGAAACACAAATGAAA 240
QY 779 ATCCCTCAGGAGTGTTCATGATCAAGCTTACTGACATCTCATGGCATGGGATCCAGG 838
DB 241 ATCCCTCAGGAGTGTTCATGATCAAGCTTACTGACATCTCATGGCATGGGATCCAGG 300
QY 839 TTCGCGTGAATGCAACAGAGTTCAACTATCTCTCTCCAGGCATCATCAACCAAAATTTGATG 898
DB 301 TTCGCGTGAATGCAACAGAGTTCAACTATCTCTCTCCAGGCATCATCAACCAAAATTTGATG 360
QY 899 CTAGATCTTGTCTGATTCATCAAGTCAAGAGAGGCTGAAATCCCTCCCAAGACCTATT 958
DB 361 CTAGATCTTGTCTGATTCATCAAGTCAAGAGAGGCTGAAATCCCTCCCAAGACCTATT 420
```


Db 721 AGAGGAAGANUTCATGATAGCCCATAGCCACCCACACAGAGTCTTACAATGAATATGCTG 780
Qy 1596 ACCCAGAGGGTGCAAGATAAATGCCATTTCCACATTTCCAGGATACACTGCGCCAGTTCAGA 1655
Db 781 CCCAGGGGCTGCAAGAACAAAGTCCCATTCACACTTTTCAGGATACGCTGCGCTAGTCCGA 840
Qy 1656 GATCTCTATTCACCAACCATCATGACTACCATCATATTTCTTCATCATCATCACAC 1707
Db 841 CGAGCTT-ATNCAACCATCATGACTACCATCATCATATTTCTGTCGACCATCATCAT 891

RESULT 13
AA013487 standard; cDNA; 551 BP.
XX AA013487;
XX 06-NOV-2001 (first entry)
XX Human LIV-1-164647 cDNA fragment (nucleotides 1690-2240).
XX Human: LIV-1-164647; cytostatic; estrogen-inducible gene; tumour;
KW cancer; breast; lung; prostate; colon; ovary; uterus; kidney; gastric;
KW salivary gland; carcinoma; drug screening; therapy; ss.
XX Homo sapiens.
XX W0200155178-A2.
XX 02-AUG-2001.
XX 25-JAN-2001; 2001W0-US02622.
XX 25-JAN-2000; 2000US-0177951.
XX 10-APR-2000; 2000US-0195761.
XX (GETH) GENENTECH INC.
XX Goddard A, Gurney AL, Smith V, Hongo JS, De Sauvage F;
XX WPI; 2001-502628/55.
XX New LIV-1 DNA and amino acids, useful for treating cancer or inhibiting
the proliferation of tumor cells in mammals, e.g. breast, lung,
prostate, colon, ovary, uterus, kidney, gastric or salivary carcinoma
-
XX Example 5; Page 76-77; 150pp; English.
XX The present sequence is human estrogen-inducible LIV-1-164647 cDNA
fragment (nucleotides 1690-2240). LIV-1 is overexpressed in tumour
tissues such as prostate, colon, lung and breast. The LIV-1 DNA and
polypeptide, and the monoclonal antibody are useful for treating cancer,
and inhibiting the proliferation of tumour cells in mammals, e.g. breast,
lung, prostate, colon, ovary, uterus, kidney, gastric or salivary
carcinoma, or other tumour cell types expressing the LIV-1-164647
protein. In particular, the mammal is a human. The LIV-1 DNA and
polypeptide may also be used in screening assays for drug candidates.
XX Sequence 551 BP; 128 A; 130 C; 127 G; 166 T; 0 other;
SQ

Query Match 24.2%; Score 549.4; DB 22; Length 551;
Best Local Similarity 99.8%; Pred. No. 7.5e-140;
Matches 550; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1618 TGGCATTCACATTTCCAGATACACTCGCGCAGTCAGACGATCTCATTCACACATCATCAT 1677
Db 1 TGGCATTCACATTTCCAGATACACTCGCGCAGTCAGACGATCTCATTCACACATCATCAT 60
Qy 1678 GACTTACCATCATATTTCTCCATCATCATCACCACCAACCAACCAACCAACCAACCAACCA 1737
Db 61 GACTTACCATCATATTTCTCCATCATCATCACCACCAACCAACCAACCAACCAACCAACCA 120

Qy 1738 CAGCGCTACTCTCGGAGGAGAGCTGAAAGATGCGGGCGTCCGCACATTTGGCCCTGGATGGTG 1797
Db 121 CAGCGCTACTCTCGGAGGAGAGCTGAAAGATGCGGGCGTCCGCACATTTGGCCCTGGATGGTG 180
Qy 1798 ATAATGGGTGATCGCCCTGCACAAATTTTCAGCGATGGCTAGCAATTCGGTCTGCTTTTACT 1857
Db 181 ATAATGGGTGATGGCCCTGCACAAATTTTCAGCGATGGCTAGCAATTCGGTCTGCTTTTACT 240
Qy 1858 GAAGGCTTATCAAGTGGTTTAAAGTACTTCTGTGCTGTGTTCTGTGCATGAGTTGGCTCAT 1917
Db 241 GAAGGCTTATCAAGTGGTTTAAAGTACTTCTGTGCTGTGTTCTGTGCATGAGTTGGCTCAT 300
Qy 1918 GAATTAGGTGACTTTGCTGTTTACTTAAAGGCTGGCATGACCGTTAAGCAGCGTGTCTCTT 1977
Db 301 GAATTAGGTGACTTTGCTGTTTACTTAAAGGCTGGCATGACCGTTAAGCAGCGTGTCTCTT 360
Qy 1978 TATAATGCATTGTCACCCATGCTGGCGTATCTTTGGAAATGCAACAGCAATTTTCATTGGT 2037
Db 361 TATAATGCATTGTCACCCATGCTGGCGTATCTTTGGAAATGCAACAGCAATTTTCATTGGT 420
Qy 2038 CATTATGCTGAAATGTTTCTTATGTGATATTTGACACTTACTGCTGCTGCTTATTCATGTAT 2097
Db 421 CATTATGCTGAAATGTTTCTTATGTGATATTTGACACTTACTGCTGCTGCTTATTCATGTAT 480
Qy 2098 GTTGCTCTGTTGATATGTGACCTGAAATGCTGCACAAATGATGCTAGTGACCATGATGT 2157
Db 481 GTTGCTCTGTTGATATGTGACCTGAAATGCTGCACAAATGATGCTAGTGACCATGATGT 540
Qy 2158 AGCCGCTGGGG 2168
Db 541 AGCCGCTGGGG 551

RESULT 14
AAZ33622
ID AAZ33622 standard; cDNA; 1597 BP.
XX AAZ33622;
XX 08-DEC-1999 (first entry)
XX Human breast tumour-associated EST 12.
XX Expressed sequence tag; EST; human; breast; cancer; gene therapy;
KW treatment; tumour; cytostatic; medicament; ss.
XX Homo sapiens.
XX DE19813839-A1.
XX 23-SEP-1999.
XX 20-MAR-1998; 98DE-1013839.
XX 20-MAR-1998; 98DE-1013839.
XX (META-) METAGEN GES GENOMFORSCHUNG MBH.
XX Specht T, Hinzmann B, Schmitt A, Pillarsky C, Dahl E, Rosentahl A;
XX WPI; 1999-528981/45.
XX Human nucleic acid sequences and protein products from tumor breast
tissue, useful for breast cancer therapy -
XX Claim 3; 93; 188pp; German.
XX This invention describes novel human nucleic acid sequences from tumor
breast tissue which have cytostatic activity. The nucleic acid sequences
can be used to produce and isolate full-length gene sequences. They can
be used to express proteins, which can be used as tools to find an
activity against breast cancer. The sequences can be used in sense or
antisense form. They are especially useful for medicaments for gene

CC therapy to treat breast cancer. AAZ33611-248617 represents expressed
CC sequence tags described in the method of the invention.

XX
SQ Sequence 1597 bp; 476 A; 229 C; 337 G; 555 T; 0 other;

Query Match 22.5%; Score 509.4; DB 20; LenqLh 1597;
Best Local Similarity 99.8%; Pred. No. 1.2e-128;
Matches 510; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1758 GCTGAAGATGCGCGCTGCCACATTTGGCTGGATGGTGATTAATGGGTGATGCA 1817
Db 19 GCTGAAGATGCGCGCTGCCACATTTGGCTGGATGGTGATTAATGGGTGATGCA 78
QY 1818 CAATTTACGATGGCTAGCAATTTGGCTGCTTTTACTGAAGCTTTATCAAGTGGT 1877
Db 79 CAATTTACGATGGCTAGCAATTTGGCTGCTTTTACTGAAGCTTTATCAAGTGGT 138
QY 1878 AGTACTTCTGTTGTTGTTCTGTCATGAGTTGGCTCATGAATTTAGTGTGCT 1937
Db 139 AAGTATTTCTGTTGTTGTTCTGTCATGAGTTGGCTCATGAATTTAGTGTGCT 198
QY 1938 TCTACTAAAGGCTGGCATGACCGTTAAGCAGCTGCTCTTTTATAATGCAATTTCTCAGCAAT 1997
Db 199 TCTACTAAAGGCTGGCATGACCGTTAAGCAGCTGCTCTTTTATAATGCAATTTCTCAGCAAT 258
QY 1998 GCTGGCTATCTTTGAATGGCAACAGGAATTTTCAATTTGCTCATTTGCTGAAAATGTTTC 2057
Db 259 GCTGGCTATCTTTGAATGGCAACAGGAATTTTCAATTTGCTCATTTGCTGAAAATGTTTC 318
QY 2058 TATGCTGATATTTGACATTTACTGCTGGCTATTTCAATGATGTTGCTGTTGATATGTT 2177
Db 319 TATGCTGATATTTGACATTTACTGCTGGCTATTTCAATGATGTTGCTGTTGATATGTT 378
QY 2118 ACCTGAAATGTCACAAATGATGCTAGTACCATGATGATAGCCCTGGGGTATTTCTT 2177
Db 379 ACCTGAAATGTCACAAATGATGCTAGTACCATGATGATAGCCCTGGGGTATTTCTT 438
QY 2178 TTTACAGATGCTGGGATGCTTTTGGGTTTGGAAATTTATGTTACTTATTTCCATATTTGA 2237
Db 439 TTTACAGATGCTGGGATGCTTTTGGGTTTGGAAATTTATGTTACTTATTTCCATATTTGA 498
QY 2238 ACATAAATCGTGTTCGTATATAATTTCTAG 2268
Db 499 ACATAAATCGTGTTCGTATATAATTTCTAG 529

RESULT 15
AAL19686/c
XX AAL19686 standard; cDNA: 424 bp.
AC AAL19686;
XX
XX
XX 07-DEC-2001 (first entry)
XX human breast cancer expressed polynucleotide 12143.
XX
XX Human; breast cancer; cell marker; cytostatic; ss.
XX Homo sapiens.
XX W0200151628-A2.
XX
XX 19-JUL-2001.
XX
XX 10-JAN-2001; 2001WO-US00798.
XX
XX 14-JAN-2000; 2000US-0176077.
PR 14-MAR-2000; 2000US-0189167.
PR 24-MAR-2000; 2000US-0192099.
PR 29-MAR-2000; 2000US-0193480.
PR 15-MAY-2000; 2000US-0205230.
PR 09-JUN-2000; 2000US-0211315.
PR 25-JUL-2000; 2000US-0220534.

XX
PA (MILL-) MILLENNIUM PREDICTIVE MEDICINE INC.
XX
PI Lillie J, Xu Y, Wang Y, Steinmann K;
XX
DR WPI: 2001-451856/48.
XX
PT New peptide useful as a marker for the diagnosis of breast cancer -
XX
PS Claim 1; Page 2147; 3695pp; English.
XX
CC The invention relates to human breast cancer expressed polynucleotides
CC (AAL07544-AAL26789) and methods of assessing whether a patient is
CC afflicted with breast cancer by examining the correlation between the
CC expression of certain markers and the cancerous state of breast cells.
CC The polynucleotides and encoded polypeptides are potential markers for
CC detecting, diagnosing, monitoring, characterizing and treating
CC potentially preventing breast cancer. The polynucleotides and encoded
CC polypeptides are also useful for isolating compounds with cytostatic
CC activity.

SQ Sequence 424 bp; 106 A; 89 C; 80 G; 149 T; 0 other;
Query Match 17.2%; Score 390.4; DB 22; Length 424;
Best Local Similarity 97.1%; Pred. No. 2.1e-96;
Matches 408; Conservative 0; Mismatches 11; Indels 1; Gaps 1;

QY 1179 ACCAGCAATGGAATGAAAGAGAGACCATTTTTCAGTCTCTCTCTCAAAACATAGA 1238
Db 420 ACCAAGCATGAAATGAAAGAGAGAGACCATTTTTCAGTCTCTCTCTCAAAACATAGA 362
QY 1239 AGAAGTGGCTATTTTGGATTTCCAGTGGAGGGTCTTAACAGCTCTAGGAGGCTGTTT 1298
Db 361 AGAAGTGGCTATTTTGGATTTCCAGTGGAGGGTCTTAACAGCTCTAGGAGGCTGTTT 302
QY 1299 CATGTTTCTTTGTTGAACATGCTCTCACATTTGATCAACCAATTTAAAGATAAGAGAAA 1358
Db 301 CATGTTTCTTTGTTGAACATGCTCTCACATTTGATCAACCAATTTAAAGATAAGAGAAA 242
QY 1359 GAATCAGAAACCTGAAATGATGATGTTGGAGATTAAAGAGAGTGTTCAGTA 1418
Db 241 GAATCAGAAACCTGAAATGATGATGTTGGAGATTAAAGAGAGTGTTCAGTA 182
QY 1419 TGAATCTCAACTTTTCAACAAATGAGGAGAAAGTAGATACAGATGATCGAACTCAAGGCTA 1478
Db 181 TGAATCTCAACTTTTCCACAAATGAGGAGAAAGTAGATACAGATGATCGAACTCAAGGCTA 122
QY 1479 TTTACGAGAGCTCACAAGAGCCCTCCCACTTTTGTATTTCTCAGAGCTGCACTTTGGA 1538
Db 121 TTTACGAGAGCTCACAAGAGCCCTCCCACTTTTGTATTTCTCAGAGCTGCACTTTGGA 62
QY 1539 AGAAGAGAGTCTATGATGATGCTCATCTCATCCACAGGAAGTCTACAATGAATATGATCC 1598
Db 61 AGAAGAGAGTCTATGATGATGCTCATCTCATCCACAGGAAGTCTACAATGAATATGATCC 2

Search completed: November 28, 2002, 02:32:04
Job time : 472 secs

Db	1818	TACCATCATATTCTCCATCATACACACACCAAAACACCACATCCTCACAGTCACAGCCAG	1877
Qy	1741	CGCTACTCTCGGGAGGAGCTGAAAGATGCCGGCTCGCCACTTTGGCCTGGATGGTGATA	1800
Db	1878	CGCTACTCTCGGAGGAGCTGAAGATGCCGGCTCGCCACTTTGGCCTGGATGGTGATA	1937
Qy	1801	ATGGGTGATGGCCTGACAAATTCAGCGATGGCTAGCAATTGGTGCTGCTTTACTGAA	1860
Db	1938	ATGGGTGATGGCCTGCACAAATTCAGCGATGGCTAGCAATTGGTGCTGCTTTACTGAA	1997
Qy	1861	GGCTATCAAGTGGTTTAAGTACTCTGTTGCTGCTGTTCTGTCATGAGTTGCCTCATGAA	1920
Db	1998	GGCTATCAAGTGGTTTAAGTACTCTGTTGCTGTTCTGTCATGAGTTGCCTCATGAA	2057
Qy	1921	TTAGGTGACTTTGCTGTTCTTACTAAAGGCTGGCATGACCGTTAAGCAGGCTGTGCTTTAT	1980
Db	2058	TTAGGTGACTTTGCTGTTCTTACTAAAGGCTGGCATGACCGTTAAGCAGGCTGTGCTTTAT	2117
Qy	1981	AATGCATTCTCAGCCATGCTGGCGTATCTTGGATGGCAACAGAGAAATTTTCATGGTCAAT	2040
Db	2118	AATGCATTCTCAGCCATGCTGGCGTATCTTGGATGGCAACAGAGAAATTTTCATGGTCAAT	2177
Qy	2041	TATGCTGAAAAATCTTCTATGTGGATATTTCACCTTACTGCTGGCTATTTCATGATGTT	2100
Db	2178	TATGCTGAAAAATGTTTCTATGTGGATATTTCACCTTACTGCTGGCTATTTCATGATGTT	2237
Qy	2101	GCTCTGGTTGATATGGTACCTGGAATGCTGCACAATGATGCTAGTGACCATGGATGAGC	2160
Db	2238	GCTCTGGTTGATATGGTACCTGGAATGCTGCACAATGATGCTAGTGACCATGGATGAGC	2297
Qy	2161	CGCTGGGGGTATTCTTTTACAGAAATCTGGGATGCTTTTGGGTTTGGAAATTAAGTTA	2220
Db	2298	CGCTGGGGGTATTCTTTTACAGAAATCTGGGATGCTTTTGGGTTTGGAAATTAAGTTA	2357
Qy	2221	CTTATTTCCATATTTGAACATAAAATCGTTTCCGTATAAATTTCTAG	2268
Db	2358	CTTATTTCCATATTTGAACATAAAATCGTTTCCGTATAAATTTCTAG	2405

RESULT 2	AX207207	2776 bp	DNA	linear	PAT 30-AUG-2001
LOCUS	AX207207				
DEFINITION	Sequence 3 from Patent WO0155178.				
ACCESSION	AX207207				
VERSION	AX207207.1	GI:15394961			
KEYWORDS	human.				
SOURCE	Homo sapiens				
ORGANISM	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.				
REFERENCE	1 (bases 1 to 2776)				
AUTHORS	Goddard, A., Gurney, A.L., Smith, V., Hongo, J.A. and de Sauvage, F.				
TITLE	Compositions and methods for treatment of cancer				
JOURNAL	Patent: WO 0155178-A 3 02-AUG-2001; GENENTECH, INC. (US)				
FEATURES	Location/Qualifiers				
source	1..2776				
	/organism="Homo sapiens"				
	/db_xref="taxon:9606"				
BASE COUNT	823 a	583 c	592 g	778 t	
ORIGIN					

	Query Match	99.9%	Score 2264.8;	DB 6;	Length 2776;
	Best Local Similarity	99.9%;	Pred. No. 0;	Mismatches 2266;	Conservative 0;
					Gaps 0;
Qy	1	ATGGCGAGGAAGTTATCTGTAATCTTGATCTGTGACCTTGCCTTCCTCTGTCACAATAATCCC			
Db	73	ATGGCGAGAAGTTATCTGTAATCTTGATCTGTGACCTTGCCTTCCTCTGTCACAATAATCCC			
Qy	61	CTTCATGAACATAAAGCAGCTGCTTTCCGCCAGACCACACTGAGAAAAATTAGTCGGAATTCGG			

Db	133	CTTCATGAACCTAAAGCAGCTGCTTTTCCCCAGACCCACTGAGAAAAATTAGTCGGAATTGG	192
Qy	121	GAATCTGGCATTAAATGTTGACTTTGGCAATTTCCACAGCGCAATATCATCTACAACAGCTT	180
Db	193	GAATCTGGCATTAAATGTTGACTTTGGCAATTTCCACAGCGCAATATCATCTACAACAGCTT	252
Qy	181	TTCTACCGCTATGGAGAAAAATAATTTCTTTGTGCTGAGTTGAAGGGTTTCAGAAAAATTACTTTCAA	240
Db	253	TTCTACCGCTATGGAGAAAAATAATTTCTTTGTGCTGAGTTGAAGGGTTTCAGAAAAATTACTTTCAA	312
Qy	241	AATATAGGCATAGATAAGATTAAAAGATCCATATACACCATGACCAGCACCACCTACTCA	300
Db	313	AATATAGGCATAGATAAGATTAAAAGATCCATATACACCATGACCAGCACCACCTACTCA	372
Qy	301	GACCACGAGCATCACTCAGACCATCAGCGTGCACTCAGACCATGAGCATCACTCAGACCAC	360
Db	373	GACCACGAGCATCACTCAGACCATGAGCGTGCACTCAGACCATGAGCATCACTCAGACCAC	432
Qy	361	GAGCATCACTCTGACCATGATCATCACTCTCACCATAATCATGCTGTTCTGGTAAAAAT	420
Db	433	GAGCATCACTCTGACCATGATCATCACTCTCACCATAATCATGCTGTTCTGGTAAAAAT	492
Qy	421	AAGCGAAAAGCTCTTTTGCCAGACCATCACTCAGATAGTTCAGSTAAGATCCTAGAAAC	480
Db	493	AAGCGAAAAGCTCTTTTGCCAGACCATCACTCAGATAGTTCAGSTAAGATCCTAGAAAC	552
Qy	481	AGCCAGGGGAAAAGGAGCTCACCGACCGAGAACATGCCAGTGGTGAAGGAATGTCAAGGAC	540
Db	553	AGCCAGGGGAAAAGGAGCTCACCGACCGAGAACATGCCAGTGGTGAAGGAATGTCAAGGAC	612
Qy	541	AGTGTTTAGTGTCTAGTGAAGTGACCTCAACTGTGTACAACATGCTCTCTGAAGGAACCTCAC	600
Db	613	AGTGTTTAGTGTCTAGTGAAGTGACCTCAACTGTGTACAACATGCTCTCTGAAGGAACCTCAC	672
Qy	601	TTTCTTAGAGACAATAGAGACTCCAGAGACTGGAAAACTCTTCCCCAAAAGATGTAAAGCAGC	660
Db	673	TTTCTTAGAGACAATAGAGACTCCAGAGACTGGAAAACTCTTCCCCAAAAGATGTAAAGCAGC	732
Qy	661	TCCACTCCACCCAGTGTCACTCAAAGAGCGGGGTGAGCCGGCTGGCTGGTGAAGAAAACA	720
Db	733	TCCACTCCACCCAGTGTCACTCAAAGAGCGGGGTGAGCCGGCTGGCTGGTGAAGAAAACA	792
Qy	721	AATGAATCTGTGAGTGAGCCCCGAAAAGGCTTTATGTATCCAGAAAACACAAATGAAAAT	780
Db	793	AATGAATCTGTGAGTGAGCCCCGAAAAGGCTTTATGTATCCAGAAAACACAAATGAAAAT	852
Qy	781	CCCTCAGGAGTGTTCATAGCATCAAGCTACTGACATCTCATGCGCATGGGCATCCAGGTT	840
Db	853	CCCTCAGGAGTGTTCATAGCATCAAGCTACTGACATCTCATGCGCATGGGCATCCAGGTT	912
Qy	841	CCGCTGAATGCAACAGAGTTCAACTATCTCTGTCAGCGCATCATCAACCAATTTGATGCT	900
Db	913	CCGCTGAATGCAACAGAGTTCAACTATCTCTGTCAGCGCATCATCAACCAATTTGATGCT	972
Qy	901	AGATCTTGCTGATTCATACAAGTGAAGAAGGCTGAAATCCCTCCAAAAGACCTATTCA	960
Db	973	AGATCTTGCTGATTCATACAAGTGAAGAAGGCTGAAATCCCTCCAAAAGACCTATTCA	1032
Qy	961	TTACAATAGCCCTGGGTGGTGGTTTTATAGCCATTTCCATCATCAGTTTCTGTGCTCTG	1020
Db	1033	TTACAATAGCCCTGGGTGGTGGTTTTATAGCCATTTCCATCATCAGTTTCTGTGCTCTG	1092
Qy	1021	CTGGGGTTATCTTAGTGCCTCTCATGAATCGGGTGGTTTTTCAAAATTTCTCCTGAGTTTC	1080
Db	1093	CTGGGGTTATCTTAGTGCCTCTCATGAATCGGGTGGTTTTTCAAAATTTCTCCTGAGTTTC	1152
Qy	1081	CTTGTGGCACTGGCCGTTGGGACITTTGAGTGGTGATGCTTTTTTACACCTTCTTCCACAT	1140
Db	1153	CTTGTGGCACTGGCCGTTGGGACITTTGAGTGGTGATGCTTTTTTACACCTTCTTCCACAT	1212
Qy	1141	TCATCTGAAGTCAACCATAGTCATAGCCATGAAGCAACAGCAATGGAATGAAAGA	1200
Db	1213	TCATCTGAAGTCAACCATAGTCATAGCCATGAAGCAACAGCAATGGAATGAAAGA	1272

QY	1201	GGACCATT	TTTCAGT	CTCATG	CTCTCTCAAAACATAGAGAAAGTGCCATTTTGGATTCC	1261
DB	1273	GGACCATT	TTTCAGT	CTCATG	CTCTCTCAAAACATAGAGAAAGTGCCATTTTGGATTCC	1332
QY	1261	ACGTGGAAGG	CTACAG	CTCTAGAG	GCCTGTATTTTCATGTTCTTGTGTTGAACATGTC	1320
DB	1333	ACGTGGAAGG	CTACAG	CTCTAGAG	GCCTGTATTTTCATGTTCTTGTGTTGAACATGTC	1392
QY	1321	CTCACAT	TGATCA	AAACAATTTAAAGAT	AAGAAGAAAAAGATCAGAGAACCCTGAAAT	1380
DB	1393	CTCACAT	TGATCA	AAACAATTTAAAGAT	AAGAAGAAAAAGATCAGAGAACCCTGAAAT	1452
QY	1381	GATGATGAT	GTGGAG	ATTAAAGACAG	TGTGCCAAGTATGAATCTCAACTTTTCAACAAT	1440
DB	1453	GATGATGAT	GTGGAG	ATTAAAGACAG	TGTGCCAAGTATGAATCTCAACTTTTCAACAAT	1512
QY	1441	GAGCAGAAAG	TAGATAC	AGATCGA	ACTCGAAGCTATTTACCAGCAGACTCACAGAG	1500
DB	1513	GAGCAGAAAG	TAGATAC	AGATCGA	ACTCGAAGCTATTTACCAGCAGACTCACAGAG	1572
QY	1501	CCCTCCCA	CTTTGAT	TTCTCAGCAG	CTCGAGTCTTGGAGAAGAGAGGTCATGATAGCT	1560
DB	1573	CCCTCCCA	CTTTGAT	TTCTCAGCAG	CTCGAGTCTTGGAGAAGAGAGGTCATGATAGCT	1632
QY	1561	CATGCTCAT	CCACAGA	AGTCTAC	AANTGAATGTACCCAGAGGGTGCAAGATAAATGC	1620
DB	1633	CATGCTCAT	CCACAGA	AGTCTAC	AANTGAATGTACCCAGAGGGTGCAAGATAAATGC	1692
QY	1621	CATTACAT	TTTCCACG	ATACACTCGG	CCAGTCAGACGATCTCATTCACCAACATCATGAC	1680
DB	1693	CATTACAT	TTTCCACG	ATACACTCGG	CCAGTCAGACGATCTCATTCACCAACATCATGAC	1752
QY	1681	TACCATCAT	TATTCAT	CATCACCA	CCACCAAAACACCATCTCACAGTCACAGCCAG	1740
DB	1753	TACCATCAT	TATTCAT	CATCACCA	CCACCAAAACACCATCTCACAGTCACAGCCAG	1812
QY	1741	CGCTACT	CTCGGG	AGGAGCTG	AAAGATGCGCGGCTCGCCACTTTGGGCTGGATGGTGATA	1800
DB	1813	CGCTACT	CTCGGG	AGGAGCTG	AAAGATGCGCGGCTCGCCACTTTGGGCTGGATGGTGATA	1872
QY	1801	ATGGGTGAT	GCGCTGC	ACAATTCAG	CGATGGCTAGCAATGGTGTGCTTTACTGAA	1860
DB	1873	ATGGGTGAT	GCGCTGC	ACAATTCAG	CGATGGCTAGCAATGGTGTGCTTTACTGAA	1932
QY	1861	GGCTTAT	CAAGTGG	TTAACTCT	TGCTGTGTTCTGTATGAGTTGGCTCATGAA	1920
DB	1933	GGCTTAT	CAAGTGG	TTAACTCT	TGCTGTGTTCTGTATGAGTTGGCTCATGAA	1992
QY	1921	TTAGTGAT	CAATTTGCT	TACTAAAG	CGCTGCTATGACCTTAAGCAGGCTGCTCTTAT	1980
DB	1993	TTAGTGAT	CAATTTGCT	TACTAAAG	CGCTGCTATGACCTTAAGCAGGCTGCTCTTAT	2052
QY	1981	AATGCAT	TTGTCAG	CCATGCTGG	CGCTATCTGGAATGCAACAGGAATTTTCATGGTGAT	2040
DB	2053	AATGCAT	TTGTCAG	CCATGCTGG	CGCTATCTGGAATGCAACAGGAATTTTCATGGTGAT	2112
QY	2041	TATGCTG	AAATGTTT	CTATGTG	GAATTTGCACTTACTGCTGGCTTATTCATGATGTT	2100
DB	2113	TATGCTG	AAATGTTT	CTATGTG	GAATTTGCACTTACTGCTGGCTTATTCATGATGTT	2172
QY	2101	GCCTGGT	TGATAT	GGTACCT	GAAATGCTCACAATGATGCTAGTGACCATGGATGATG	2160
DB	2173	GCCTGGT	TGATAT	GGTACCT	GAAATGCTCACAATGATGCTAGTGACCATGGATGATG	2232
QY	2161	CGCTGGG	GGTAT	TTCTTTT	TACAAATGCTGGGATGCTTTTGGGTTTGGAAATATGTTA	2220
DB	2233	CGCTGGG	GGTAT	TTCTTTT	TACAAATGCTGGGATGCTTTTGGGTTTGGAAATATGTTA	2292
QY	2221	CTTATTT	CCATAT	TTGAACA	TAAATCGTGTTCGTATAAATTTCTAG	2268
DB	2293	CTTATTT	CCATAT	TTGAACA	TAAATCGTGTTCGTATAAATTTCTAG	2340

RESULT 3	
HSU41060	
LOCUS	
DEFINITION	
ACCESSION	
VERSION	
KEYWORDS	
SOURCE	
ORGANISM	
REFERENCE	
AUTHORS	
TITLE	
JOURNAL	
REFERENCE	
AUTHORS	
TITLE	
JOURNAL	
COMMENT	
FEATURES	
SOURCE	

HSU41060 2744 bp mRNA linear PRI 08-FEB-2001
Homo sapiens estrogen regulated LIV-1 protein (LIV-1) mRNA,
complete cds.
U41060
U41060.2 GI:12717792

KEYWORDS	.
SOURCE	Homo sapiens.
ORGANISM	Homo sapiens
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
AUTHORS	1 (bases 1 to 2744) Green.C. and Morgan.H.
TITLE	Direct Submission
JOURNAL	Submitted (08-FEB-2001) Biochemistry, University of Liverpool, P.O. Box 147, Liverpool L69 3BX, UK
REFERENCE	2 (bases 1 to 2744) Green.C., Gilhooly.E.M. and Walker.N.J.
AUTHORS	Direct Submission
TITLE	Submitted (21-NOV-1995) Biochemistry, University of Liverpool, P.O. Box 147, Liverpool L69 3BX, UK
JOURNAL	On Feb 8, 2001 this sequence version replaced gi:1256000.
COMMENT	

```

source
1. .2744
/organism="Homo sapiens"
/db_xref="taxon:9606"
/cell_line="MCF-7 human breast cancer cell line"
/notes="estrogen induced mRNA"
1. .2744
/gene="LIV-1"
138. .2387
/gene="LIV-1"
/notes="breast cancer"
/codon_start=1
/product="estrogen regulated LIV-1 protein"
/protein_id="AAA96258.2"
/db_xref="GI:12711793"

```

/translation="MARKLSVLILITFALSVTNPNLHELKAAAFQOTTEKISPNWESG
NVDLAISTROYHLQOLFYRYKGNQDQNLQNGIDKIIRIHIDHDHSDH
EHUSDHERSDHEHSDHSDHHAAGNKKALCPDHDSSDQKPRNSQKGA
HREPHASGRNNKQSVASEVSTVNTVSECTHELETETPRPKLFPKXDSVSTP
SVTSKRSVLAKRGTNYESPRKGYMSRTNENPOECENASKLLTSHCMGIOVPL
NATEFINCPAIINGIDARSCILHTSEKKAETLPKPYSLQIATWVGFIATISIFLSL
LGVILYPLMNRVFFELFSLVALAVGTLTSGDAFLHLLPHSHASHHHSHSHEPAMEM
KRGPLESHLSSONIESAYEDSTWGLGTALGGLFMFLVHVHLTKOFKDKKKKQK
KPNDDOEHKQLSQSLEENKVDOTDRTEGYLRADSPEDSHSDSQOAVLEE
EYMIADAHPEQYNEYSPKCKNKHSHFDITGLQSDLLHHHDYHHLIHHHHHOM
HHPSHSQRYSREELKAGVTAWMIWIMGDLHNFSODGLAIGAAFTEGLSGLSTSV
AFVCEHPELHGDLQAVLLKAGMTQAQVLYNALSNMLAYLGMATGTEIGHYANVSH
IFALTAGLVEYVALVDMPPEMLHNDASDHGSGSRGQYFFLQNGAGLGLFGMLLISL
HKTWRDINE"

BASE COUNT			HKLIVFRINE [®]		
	786 a	585 c	601 q	772 t	
1	100	100	100	100	100
2	100	100	100	100	100
3	100	100	100	100	100
4	100	100	100	100	100
5	100	100	100	100	100
6	100	100	100	100	100
7	100	100	100	100	100
8	100	100	100	100	100
9	100	100	100	100	100
10	100	100	100	100	100
11	100	100	100	100	100
12	100	100	100	100	100
13	100	100	100	100	100
14	100	100	100	100	100
15	100	100	100	100	100
16	100	100	100	100	100
17	100	100	100	100	100
18	100	100	100	100	100
19	100	100	100	100	100
20	100	100	100	100	100
21	100	100	100	100	100
22	100	100	100	100	100
23	100	100	100	100	100
24	100	100	100	100	100
25	100	100	100	100	100
26	100	100	100	100	100
27	100	100	100	100	100
28	100	100	100	100	100
29	100	100	100	100	100
30	100	100	100	100	100
31	100	100	100	100	100
32	100	100	100	100	100
33	100	100	100	100	100
34	100	100	100	100	100
35	100	100	100	100	100
36	100	100	100	100	100
37	100	100	100	100	100
38	100	100	100	100	100
39	100	100	100	100	100
40	100	100	100	100	100
41	100	100	100	100	100
42	100	100	100	100	100
43	100	100	100	100	100
44	100	100	100	100	100
45	100	100	100	100	100
46	100	100	100	100	100
47	100	100	100	100	100
48	100	100	100	100	100
49	100	100	100	100	100
50	100	100	100	100	100
51	100	100	100	100	100
52	100	100	100	100	100
53	100	100	100	100	100
54	100	100	100	100	100
55	100	100	100	100	100
56	100	100	100	100	100
57	100	100	100	100	100
58	100	100	100	100	100
59	100	100	100	100	100
60	100	100	100	100	100
61	100	100	100	100	100
62	100	100	100	100	100
63	100	100	100	100	100
64	100	100	100	100	100
65	100	100	100	100	100
66	100	100	100	100	100
67	100	100	100		

Query Match	98.0%	Score 2222;	DB 9;	Length 2744;
Best Local Similarity	99.2%;	Pred. No. 0;		
Matches 2250;	Conservative	0;	Mismatches 0;	Indels 18;
				Gaps 1;

1	ATGCGAGGAGTTATCTGTAATCTTGATCCTGACCTTTGGCCCTCTCTGTACAAATCCC	60
138	ATGCGAGGAGTTATCTGTAATCTTGATCCTGACCTTTGGCCCTCTCTGTACAAATCCC	197
61	CTTCATGAACCTAAAGCAGCTGCTTTCCCCAGACACCTGAGAAAATTAGTCGGAATGG	120
198	CTTCATGAACCTAAAGCAGCTGCTTTCCCCAGACACCTGAGAAAATTAGTCGGAATGG	257
121	GAATCTGGCATTAATGTTGACTTTGGCAATTTCCACACGGCAATATCATCTACACACGTT	180
258	GAATCTGGCATTAATGTTGACTTTGGCAATTTCCACACGGCAATATCATCTACACACGTT	317
181	TTTACCGCTATGGAGAAAATAATCTTTGTCAAGGAGGTTGAGAAAATTAATCTTCAA	240

GenCore version 5.1.3
Copyright (c) 1993 - 2002 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: November 27, 2002, 22:28:40 : Search time 5666 Seconds
(without alignments)
11649.341 Million cell updates/sec

Title: US-09-642-034-4
Perfect score: 2268
Sequence: 1 atggcgaggaagtattatctgt.....tggttcgtataataattttctag 2268

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 2054640 seqs, 14551402878 residues
Total number of hits satisfying chosen parameters: 4109280

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : GenEmbl.*

- 1: gb.ba.*
- 2: gb.htg.*
- 3: gb.in.*
- 4: gb.om.*
- 5: gb.ov.*
- 6: gb.pat.*
- 7: gb.ph.*
- 8: gb.pl.*
- 9: gb.pr.*
- 10: gb.ro.*
- 11: gb.sts.*
- 12: gb.sy.*
- 13: gb.un.*
- 14: gb.vi.*
- 15: em.ba.*
- 16: em.fun.*
- 17: em.hum.*
- 18: em.in.*
- 19: em.mu.*
- 20: em.om.*
- 21: em.or.*
- 22: em.ov.*
- 23: em.pat.*
- 24: em.ph.*
- 25: em.pl.*
- 26: em.ro.*
- 27: em.sts.*
- 28: em.un.*
- 29: em.vi.*
- 30: em.htg_hum.*
- 31: em.htg_inv.*
- 32: em.htg_other.*
- 33: em.htg_mus.*
- 34: em.htg_pln.*
- 35: em.htg_rod.*
- 36: em.htg_mam.*
- 37: em.htg_vrt.*
- 38: em_sy.*
- 39: em_htgo_hum.*
- 40: em_htgo_mus.*
- 41: em_htgo_other.*

Pred. No. is the number of results predicted by chance to have a

score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Match	Length	DB	ID	Description
1	2268	100.0	3461	6	AX465588	Sequence
2	2264.8	99.9	2776	6	AX207207	Sequence
3	2222	98.0	2744	9	HS041060	U41060 Homo sapien
4	2198	96.9	3461	6	AX207205	Sequence
5	1616	71.3	3287	10	AB071697	AB071697 Mus muscu
6	1263	55.7	1310	6	I76891	Sequence 1
7	1263	55.7	2404	6	I76892	Sequence 3
8	1151	50.7	2236	9	BC008317	BC008317 Homo sapi
9	789	34.8	160170	2	AP001158	AP001158 Homo sapi
10	787.4	34.7	221941	9	AC091060	AC091060 Homo sapi
11	787.4	34.7	224788	2	AP001905	AP001905 Homo sapi
12	619.6	27.3	899	6	AX465590	Sequence
13	549.4	24.2	551	6	AX207216	Sequence
14	509.4	22.5	1597	6	AX017261	Sequence
15	421.4	18.6	193437	2	AC126667	AC126667 Rattus no
16	319.2	14.1	337	6	AX393861	AX393861 Sequence
17	319	14.1	5231	9	AB033091	AB033091 Homo sapi
18	319	14.1	5322	6	AX405756	AX405756 Sequence
19	273.4	12.1	61103	2	AC131276	AC131276 Homo sapi
20	191.2	8.4	193437	2	AC126667	AC126667 Rattus no
21	144	6.3	52884	2	AC014796	AC014796 Drosophil
22	144	6.3	174712	3	AC010041	AC010041 Drosophil
23	144	6.3	195148	3	AC010019	AC010019 Drosophil
24	144	6.3	282267	3	AE003555	AE003555 Drosophil
25	129.2	5.7	2174	10	BC021530	BC021530 Mus muscu
26	128.4	5.7	2015	9	AB051127	AB051127 Macaca fa
27	127.6	5.6	4573	6	AX409002	AX409002 Sequence
28	127.6	5.6	4573	9	HMORFKG1P	D31887 Human mRNA
29	126	5.6	4610	6	AX078294	AX078294 Sequence
30	126	5.6	4698	6	AX281784	AX281784 Sequence
31	123.4	5.4	1383	6	AX118897	AX118897 Sequence
32	123.4	5.4	1803	9	AF193052	AF193052 Homo sapi
33	123.4	5.4	2604	9	AC027652	AC027652 Homo sapi
34	123.4	5.4	3169	9	BC012125	BC012125 Homo sapi
35	123.4	5.4	3246	9	AB040120	AB040120 Homo sapi
36	122.6	5.4	1895	9	BC027884	BC027884 Homo sapi
37	119.2	5.3	1755	9	BC015770	BC015770 Homo sapi
38	119.2	5.3	55577	2	AC014109	AC014109 Drosophil
39	119.2	5.3	170071	3	AC093497	AC093497 Drosophil
40	119.2	5.3	194979	3	AC091209	AC091209 Drosophil
41	119.2	5.3	295377	3	AE003532	AE003532 Drosophil
42	118	5.2	150	9	HUMDSF04M5	D17089 Human HapG2
43	116.2	5.1	1931	10	BC028990	BC028990 Mus muscu
44	115.8	5.1	3060	10	BC006731	BC006731 Mus muscu
45	110.6	4.9	1888	6	AX061633	AX061633 Sequence

ALIGNMENTS

RESULT 1
AX465588
LOCUS AX465588 3461 bp DNA linear PAT 16-JUL-2002
DEFINITION Sequence 1 from Patent WO0216939.
ACCESSION AX465588
VERSION AX465588.1 GI:21899891
KEYWORDS human.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1
AUTHORS Mack,D., Gish,K.C. and Wilson,K.E.
TITLE Methods of diagnosis of cancer and screening for cancer modulators
JOURNAL Patent: WO 0216939-A 1 28-FEB-2002;